

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

Volume V

1945

917 FIFTEENTH ST., N. W., • WASHINGTON 5, D. C.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT

EDITOR

G. FREDERIC KUDER The Office of the Secretary of War

ASSOCIATE EDITORS

DOROTHY C. ADKINS United States Civil Service Commission
FORREST A. KINGSBURY University of Chicago
FRED MCKINNEY, Editorial Representative of the American College
Personnel Association University of Missouri
M. W. RICHARDSON Adjutant General's Office A.U.S.

BOARD OF COOPERATING EDITORS

JOHN G. DARLEY
University of Minnesota

DAVID SEGEL
U. S. Office of Education

HAROLD A. EDGERTON
Ohio State University

C. L. SHARTLE
Ohio State University

MAX D. ENGELHART
Chicago City Junior Colleges

H. C. TAYLOR
*The W. E. Upjohn Institute
for Community Research*

E. B. GREENE
War Manpower Commission

THELMA G. THURSTONE
Chicago Teachers College

J. P. GUILFORD
University of Southern California

HERBERT A. TOOPS
Ohio State University

E. F. LINDQUIST
State University of Iowa

E. G. WILLIAMSON
University of Minnesota

CHARLES I. MOSIER
Social Security Board

BEN D. WOOD
Columbia University

P. J. RULON
Harvard University

JOHN R. YALE
Science Research Associates

The journal is open to (1) reports of research on the development and use of tests and measurements in education, government, and industry, (2) descriptions of testing programs being used for various purposes, (3) discussions of problems of measurement in general or in specific fields and (4) miscellaneous notes pertinent to the measurement field, such as suggestions of new types of items or improved methods of treating test data. Contributors receive one hundred reprints of their articles without charge. Manuscripts should be sent to G. Frederic Kuder, 415 North Oxford Street, Arlington, Virginia.

EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT is published quarterly, one volume per calendar year, at N Queen St. and McGovern Ave., Lancaster, Pa. and 217 Fifteenth St., N.W., Washington 5, D.C. Entered as second class matter October 8, 1945 at the Post Office at Lancaster, Pa., under the act of March 3, 1879.

Subscription rate \$4.00 a year, domestic and foreign. Single copies, \$1.25. Back volumes, \$5.00. Subscriptions and communications concerning change of address should be sent to 917 Fifteenth St., N.W., Washington 5, D.C.

INDEX FOR VOLUME V

<i>Berg, Irwin A. (with Graham Johnson and Robert P. Larsen)</i>	
THE USE OF AN OBJECTIVE TEST IN PREDICTING RHETORIC GRADES	429
<i>Cattell, Raymond B.</i>	
PERSONALITY TRAITS ASSOCIATED WITH ABILITIES. I. WITH INTELLIGENCE AND DRAWING ABILITIES	131
<i>Carroll, John B. (with William M. Davidson)</i>	
SPEED AND LEVEL COMPONENTS IN TIME-LIMIT SCORES: A FACTOR ANALYSIS	411
<i>Committee on Education, Society for Public Administration</i>	
THE DUTIES OF CIVIL SERVICE EXAMINERS AND TEST TECHNICIANS	371
<i>Davidson, William M. (with John B. Carroll)</i>	
SPEED AND LEVEL COMPONENTS IN TIME-LIMIT SCORES: A FACTOR ANALYSIS	411
<i>Durost, W. N. (with W. C. Kvaraceus and R. F. McClellan)</i>	
CIVILIAN TESTING IN THE QUARTERMASTER CORPS	17
<i>Dyer, Henry S.</i>	
EVIDENCE ON THE VALIDITY OF THE ARMED FORCES INSTITUTE TESTS OF GENERAL EDUCATIONAL DEVELOPMENT	321
<i>Findley, Warren G.</i>	
A GROUP TESTING PROGRAM FOR THE MODERN SCHOOL	173
<i>Fiske, Marjorie (with Paul F. Lazarsfeld)</i>	
THE OFFICE OF RADIO RESEARCH: A DIVISION OF THE BUREAU OF APPLIED SOCIAL RESEARCH	1351
<i>Franklin, Joseph Charles</i>	
DISCRIMINATIVE VALUE AND PATTERNS OF THE WECHSLER-BELLEVUE SCALES IN THE EXAMINATION OF DELINQUENT NEGRO BOYS	71
<i>Harrell, Margaret S. (with Thomas W. Harrell)</i>	
ARMY GENERAL CLASSIFICATION TEST SCORES FOR CIVILIAN OCCUPATIONS	229
<i>Harrell, Thomas W. (with Margaret S. Harrell)</i>	
ARMY GENERAL CLASSIFICATION TEST SCORES FOR CIVILIAN OCCUPATIONS	229
<i>Hartson, L. D.</i>	
INFLUENCE OF LEVEL OF MOTIVATION ON THE VALIDITY OF INTELLIGENCE TESTS	273
<i>Humm, Doncaster G.</i>	
THE RATIONALE OF TEMPERAMENT TESTING	383
<i>Jenkins, William Leroy</i>	
A QUICK GRAPHIC METHOD FOR PREDICTING M...	

<i>Johnson, Graham (with Irwin A. Berg and Robert P. Larsen)</i>	
THE USE OF AN OBJECTIVE TEST IN PREDICTING RHETORIC GRADES	429
<i>Kornhauser, Arthur</i>	
REPLIES OF PSYCHOLOGISTS TO A SHORT QUESTIONNAIRE ON MENTAL TEST DEVELOPMENTS, PERSONALITY INVENTORIES, AND THE RORSCHACH TEST ..	3
<i>Kornhauser, Arthur</i>	
REPLIES OF PSYCHOLOGISTS TO SEVERAL QUESTIONS ON THE PRACTICAL VALUE OF INTELLIGENCE TESTS	181
<i>Kvaraceus, W. C. (with W. N. Durost and R. F. McClellan)</i>	
CIVILIAN TESTING IN THE QUARTERMASTER CORPS	17
<i>Larsen, Arthur H. (with Stanley S. Marzolf)</i>	
STATISTICAL INTERPRETATION OF SYMPTOMS ILLUSTRATED WITH A FACTOR ANALYSIS OF PROBLEM CHECK LIST ITEMS	285
<i>Larsen, Robert P. (with Irwin A. Berg and Graham Johnson)</i>	
THE USE OF AN OBJECTIVE TEST IN PREDICTING RHETORIC GRADES	429
<i>Lazarsfeld, Paul F. (with Marjorie Fiske)</i>	
THE OFFICE OF RADIO RESEARCH: A DIVISION OF THE BUREAU OF APPLIED SOCIAL RESEARCH	351
<i>Lovell, Constance</i>	
A STUDY OF THE FACTOR STRUCTURE OF THIRTEEN PERSONALITY VARIABLES	335✓
<i>Mandel, Milton</i>	
TESTING FOR ADMINISTRATIVE AND SUPERVISORY POSITIONS	217
<i>Marzolf, Stanley S. (with Arthur H. Larsen)</i>	
STATISTICAL INTERPRETATION OF SYMPTOMS ILLUSTRATED WITH A FACTOR ANALYSIS OF PROBLEM CHECK LIST ITEMS	285
<i>McClellan, R. F. (with W. C. Kvaraceus and W. N. Durost)</i>	
CIVILIAN TESTING IN THE QUARTERMASTER CORPS	17
<i>Mosier, Charles I. (with M. Claire Myers and Helen G. Price)</i>	
SUGGESTIONS FOR THE CONSTRUCTION OF MULTIPLE CHOICE TESTS	261
<i>Mosier, Charles I. (with Helen G. Price)</i>	
THE ARRANGEMENT OF CHOICES IN MULTIPLE CHOICE QUESTIONS AND A SCHEME FOR RANDOMIZING CHOICES	379
<i>Myers, M. Claire (with Charles I. Mosier and Helen G. Price)</i>	
SUGGESTIONS FOR THE CONSTRUCTION OF MULTIPLE CHOICE TESTS	261
<i>Price, Helen G. (with Charles I. Mosier and M. Claire Myers)</i>	
SUGGESTIONS FOR THE CONSTRUCTION OF MULTIPLE CHOICE TESTS	261
<i>Price, Helen G. (with Charles I. Mosier)</i>	
THE ARRANGEMENT OF CHOICES IN MULTIPLE CHOICE QUESTIONS AND A SCHEME FOR RANDOMIZING CHOICES	379
<i>Schneidler, Gwendolen (with Carlos E. Ward)</i>	
THE COUNSELING PROGRAM OF THE VETERANS ADMINISTRATION	125
<i>Staff, Division of Occupational Analysis, War Manpower Commission</i>	
FACTOR ANALYSIS OF OCCUPATIONAL APTITUDE TESTS	147

<i>Strong, Jr., Edward K.</i>	
THE INTERESTS OF FOREST SERVICE MEN	157
<i>Thelen, Herbert A.</i>	
TESTING BY MEANS OF FILM SLIDES WITH SYNCHRONIZED RECORDED SOUND	33
<i>Toops, Herbert A.</i>	
SOME CONCEPTS OF JOB FAMILIES AND THEIR IMPORTANCE IN PLACEMENT	195
<i>Toops, Herbert A.</i>	
PHILOSOPHY AND PRACTICE OF PERSONNEL SELECTION	95
<i>Triggs, Frances Oralind</i>	
THE ROLE OF TESTS IN THE DIAGNOSIS AND CORRECTION OF SPELLING DEFICIENCIES OF COLLEGE STUDENTS	59
<i>Tyler, F. T</i>	
ANALYSIS OF THE TERMAN-McNEMAR TESTS OF MENTAL ABILITY	49
<i>Ward, Carlos E. (with Gwendolen Schneider)</i>	
THE COUNSELING PROGRAM OF THE VETERANS ADMINISTRATION	125
<i>Wittenborn, J. R.</i>	
MECHANICAL ABILITY, ITS NATURE AND MEASUREMENT. I. AN ANALYSIS OF THE VARIABLES EMPLOYED IN THE PRELIMINARY MINNESOTA EXPERIMENT	241
<i>Wittenborn, J. R.</i>	
MECHANICAL ABILITY, ITS NATURE AND MEASUREMENT. II. MANUAL DEXTERITY	395

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME FIVE, NUMBER ONE, SPRING

<i>Replies of Psychologists to a Short Questionnaire on Mental Test Developments, Personality Inventories, and the Rorschach Test.</i> ARTHUR KORNHAUSER	3
<i>Civilian Testing in the Quartermaster Corps.</i> W. C. KVARACEUS, W. N. DUROST, and R. F. McCLELLAN	17
<i>Testing by Means of Film Slides with Synchronized Recorded Sound.</i> HERBERT A. THELEN	33
<i>Analysis of the Terman-McNemar Tests of Mental Ability.</i> F. T. TYLER	49
<i>The Role of Tests in the Diagnosis and Correction of Spelling Deficiencies of College Students.</i> FRANCES ORALIND TRIGGS	59
<i>Discriminative Value and Patterns of the Wechsler-Bellevue Scales in the Examination of Delinquent Negro Boys.</i> JOSEPH CHARLES FRANKLIN	71
<i>Measurement Abstracts</i>	87

PRINTED IN THE UNITED STATES OF AMERICA
THE SCIENCE PRESS PRINTING COMPANY
LANCASTER, PENNSYLVANIA

REPLIES OF PSYCHOLOGISTS TO A SHORT QUESTIONNAIRE ON MENTAL TEST DEVELOPMENTS, PERSONALITY INVENTORIES, AND THE RORSCHACH TEST

ARTHUR KORNHAUSER

Bureau of Applied Social Research, Columbia University

IN APRIL 1944 a brief questionnaire was sent to 85 selected specialists on mental tests whose views might be considered representative of highly competent thought in that field. Seventy-nine completed blanks were returned (93 per cent).

The first part of the question-blank asked a few questions about the practical value of intelligence tests. This material will be summarized in a later report in this journal as well as in more popular form elsewhere.¹

The second half of the questionnaire, to be reported here, contained somewhat more technical questions. These were intended for report within the profession only, not for the general public.

Both the formulation of the questions and the selection of the expert panel were based upon personal conferences with six specialists who served as advisors. The final list of experts represents the pooled judgment of these advisors.²

¹ The set of questions on intelligence testing constituted one of a series of "polls of experts" which aim to ascertain and report to the public the conclusions of a cross-section of leading authorities on questions in their special fields. A continuing project of this kind, it is believed, may help reduce the lag of public thinking behind the views of the well-informed. The polls are intended for prompt publication in a mass-circulation magazine. While publication arrangements have been delayed during the initial stages, plans are now completed for having the poll reports appear monthly in *The American Magazine*.

² The cooperation of the mental test authorities who participated is gratefully acknowledged. In addition to those in the following list, three others requested that their names be not listed.

Dr. Dorothy C. Adkins
Dr. Anne Anastasi
Dr. Rose G. Anderson
Dr. Grace Arthur

Dr. Albert K. Kurtz
Dr. E. F. Lindquist
Dr. Irving Lorge
Comdr. C. M. Louttit

A summary of responses to the second set of questions follows.

Question 1

In the further development of mental ability testing for practical use in schools and in business, do you think most will be accomplished if psychologists concentrate on measuring separate intellectual factors or if they continue to emphasize the measurement of "general" intelligence?

	<i>No. of replies.</i>
Separate factors	55
General intelligence	5
Both "separate" and "general" checked	7
Other answer or no answer	12
	<hr/> 79

In addition to the seven who checked both, there are 9 others whose *comments* suggest that both types of emphasis are desirable. Even with these included, an overwhelming

Major Roger M. Bellows
 Dr. George K. Bennett
 Lt. Col. Robert G. Bernreuter
 Dr. Marion A. Bills
 Dr. Walter V. Bingham
 Dr. Robert A. Brotemarkle
 Dr. Andrew W. Brown
 Dr. Harold E. Burt
 Dr. Herbert S. Conrad
 Capt. Clyde H. Coombs
 Dr. Albert B. Crawford
 Dr. Edward E. Cureton
 Dr. John G. Darley
 Dr. Walter F. Dearborn
 Dr. Edgar A. Doll
 Lt. Comdr. Jack W. Dunlap
 Dr. Beatrice J. Dvorak
 Dr. Harold A. Edgerton
 Dr. Max D. Englehart
 Dr. Alvin C. Eurich
 Dr. Warren G. Findley
 Dr. Frank N. Freeman
 Dr. Douglas H. Fryer
 Dr. Henry E. Garrett
 Lt. Col. J. Guilford
 Dr. Harold O. Gulliksen
 Dr. V. A. C. Henmon
 Dr. Edwin R. Henry
 Dr. Gertrude Hildreth
 Dr. Karl J. Holzinger
 Dr. Carl Hovland
 Comdr. John G. Jenkins
 Dr. H. M. Johnson
 Dr. Truman Kelley
 Dr. G. Frederic Kuder

Lt. Col. Arthur W. Melton
 Dr. Arthur S. Otis
 Dr. Jay L. Otis
 Dr. Donald G. Paterson
 Dr. H. H. Remmers
 Lt. Col. M. W. Richardson
 Lt. Col. Robert T. Rock
 Dr. Carl R. Rogers
 Dr. Phillip Rulon
 Capt. Edward A. Rundquist
 Dr. David Segel
 Lt. Col. Morton A. Seidenfeld
 Lt. Col. Laurance F. Shaffer
 Dr. C. Shartle
 Dr. John M. Stalnaker
 Dr. Ruth Strang
 Dr. Percival M. Symonds
 Dr. E. L. Thorndike
 Capt. R. L. Thorndike
 Dr. L. L. Thurstone
 Dr. Joseph Tiffin
 Dr. Lewis E. Terman
 Dr. Herbert A. Toops
 Dr. M. R. Trabue
 Dr. Arthur E. Traxler
 Dr. Ralph W. Tyler
 Dr. Morris S. Viteles
 Dr. David Wechsler
 Dr. Frederic Lyman Wells
 Dr. Edmund G. Williamson
 Dr. Ben D. Wood
 Dr. Herbert Woodrow
 Dr. Wayne Wrightstone

majority of the 79 experts answer in favor of "separate factors."

Some typical comments from those who say both "separate factors" and "general intelligence":

Both are important. Developmental work in "separate factors" should probably be given higher priority.

"Separate factors" for a period until we find what it's all about. I don't believe tests of these "separate factors" will ever supplant entirely the test of general ability.

I think that there is still a definite use for the "general intelligence" test, but the measurement of "separate factors" should also be made.

Several others simply say, "Both needed." One interesting belief in both is expressed in these words: "'General intelligence' tests for children and group factor tests for adults." Another suggests that it is a question of the time available for testing: If only one hour, a general intelligence test; if three or four hours, tests for separate factors.

Five of the psychologists explicitly reject "general intelligence" measurement altogether. They say, for example:

"General intelligence" is a hodge-podge of several relatively independent group factors.

The concept of "general intelligence" should be entirely discarded.

"General intelligence" is like what Henry Ford said about history.

In advocating major attention to separate functions, two replies particularly stress testing for the three abilities—verbal, numerical and spatial or mechanical; seven point to the special usefulness of separate ability tests for industrial or other specific purposes; three dissociate their belief in measuring separate abilities from any particular "factor analysis" methods or any particular classification of abilities.

On behalf of continued emphasis on the measurement of "general intelligence" the following comments are made:

The value of "general intelligence" tests has been demonstrated as is indicated in their widespread use particularly in schools; the validity and value of measures of "separate factors" have yet to be shown.

With criteria as undependable as they now are, tests of general ability do as much as can be expected and tests of separate factors represent excessive refinements.

In respect to measuring isolated functions, attention is called, in four or five answers, to the limitations of this procedure. For example:

All so-called mental abilities seem to be intercorrelated. Therefore you cannot test an isolated factor if you try.

The idea of test purity is far from clear. At best factorial composition always involves a "potpourri" factor.

"Separate factors" in the strict sense cannot be now measured, but approximations can be made.

A point of some interest with respect to the tabulated replies to this question is their relationship to the age of the respondents.

The median age of those who answered "general intelligence" either alone or together with "separate factors" is 56. The median age of all others is 42. Looking at the matter in the other direction, of the 28 persons 50 years of age and over, 29 per cent answered "general intelligence"; of the 44 persons under 50, only 9 per cent answered in this way. It almost looks as though "general intelligence" is becoming an old man's concept!

Question 2

In the field of personality testing, how satisfactory or helpful for present practical use do you consider:

- (a) Personality inventories and questionnaires (such as those of Bernreuter, Bell, Humm-Wadsworth, etc.)?
- (b) The Rorschach test?

	Question (a)		Question (b)	
	No.	%	No.	%
Highly satisfactory	1	1.5	0	0.0
Moderately satisfactory	9	13.5	12	20.0
Doubtfully satisfactory	24	36.0	17	29.0
Rather unsatisfactory	22	33.0	13	22.0
Highly unsatisfactory	11	16.0	17	29.0
	67	100.0	59	100.0
Qualified, unclassifiable*	8		5	
No answer or don't know	4		15	

* Of the 8 qualified responses on question (2a), 5 tend to be favorable, 1 unfavorable, 2 neutral. Of the 5 qualified replies to question (2b), 3 are favorable and 2 unfavorable.

The ratings of the Rorschach test tend to give a flatter distribution than the other. A slightly higher percentage of the psychologists consider the Rorschach "moderately satisfactory" but a markedly higher percentage also rate it as "highly unsatisfactory." There is almost no correlation between the ratings assigned by the respondents to personality inventories and to the Rorschach tests ($r = .11$).

When the respondents are classified into clinical and non-clinical (according to their statements about their own principal types of work), some interesting differences appear in the responses to the above questions.

<i>Personality inventories</i>		<i>Clinical</i>	<i>Non-clinical</i>
Rated	Highly or moderately satisfactory	21.0%	11.0%
	Doubtfully satisfactory	39.5	34.0
	Highly or rather unsatisfactory	39.5	55.0
Total per cent		100.0% (N = 28)	100.0% (N = 38)
<i>Rorschach Test</i>		<i>Clinical</i>	<i>Non-clinical</i>
Rated:	Highly or moderately satisfactory	38.0%	11.0%
	Doubtfully satisfactory	29.0	30.0
	Highly or rather unsatisfactory	33.0	59.0
Total per cent		100.0% (N = 21)	100.0% (N = 37)

It is clear that the clinical psychologists are somewhat more favorable toward both types of tests. Their opinions differ from those of the non-clinicians particularly in respect to the Rorschach.

Similar tabulations have been made comparing the psychologists who state that a principal part of their work has dealt with personality tests and those whose work has not been principally in this field. The results are as follows:

<i>Personality Inventories</i>		<i>Work with Personality Tests</i>	
		<i>Yes</i>	<i>No</i>
	Highly or moderate satisfactory	30%	9%
	Doubtfully satisfactory	33	40
	Highly or rather unsatisfactory	37	51
		(N = 30)	(N = 35)
<i>Rorschach Test</i>		<i>Work with Personality Tests</i>	
		<i>Yes</i>	<i>No</i>
	Highly or moderately satisfactory	25%	21%
	Doubtfully satisfactory	21	32
	Highly or rather unsatisfactory	54	47
		(N = 24)	(N = 34)

Comments Regarding Personality Inventories

The most frequent comments on question 2a are those which point to the clinical and qualitative value of the blanks as contrasted with their quantitative use for purposes such as selection of personnel. Under this heading come such remarks as the following:

Useful in locating foci for further exploration and counseling, i e., qualitatively rather than quantitatively.

No use for selection, possibly some clinical value.

I believe that the personality test used as a clinical tool or used under conditions where there is full cooperation of the testee and the tester can be very illuminating and practically useful. We have not found that they give us the correct information when used as a selection tool under ordinary circumstances.

Moderately satisfactory for clinicians; highly unsatisfactory for industry. For industry, subject will not "come clean."

Most useful in counselling as points of departure, securing interest on adjustment problems, indicating students who should be referred to a psychiatrist, etc.

A second group of comments has to do with the need for validation and further research:

None of these tests has, in my opinion, been adequately validated against satisfactory outside criteria.

Highly unsatisfactory due primarily to inadequate standardization rather than to intrinsic lack of validity.

It is impossible to measure validly the gradations of personality adjustments by pencil-and-paper tests. Coarse distinctions that are reflected are usually obvious without testing. It would be helpful if psychologists as a group would go on record to this effect.

These tests validated for specific jobs have been helpful. In their present form, using standard norms, they are not helpful in industry.

Two and one-half years of work in the service have shown no helpfulness in any of these "ready-made" tests which are presumed to measure "traits" which clinicians presumed to exist. On the other hand, both Army and Navy have been able to demonstrate the use of "personality tests" constructed to meet the specific behavioral requirements of specific fields.

Other comments call attention to the value of the personality blanks insofar as they are competently and cautiously interpreted. Examples:

If used by qualified person. Should never be depended on for individual diagnosis without careful check. Of some value for research.

A very great deal depends on the skill and judgment of the person making use of the results.

In the hands of competent clinicians such devices appear quite useful—used with other data. For most counselors, who may not add salt, better counselling in regard to personality problems will be produced *without* such inventories and questionnaires.

Several replies also call attention to the differences in value among the various inventories in use. As one reply puts it:

There are some 500 personality tests, most of which are of little or no value as measurement devices. A few, probably not more than a dozen, could be recommended for experimental use in a testing program.

In the few comments regarding particular blanks, the Bell and the Minnesota multiphasic inventories receive favorable mention, the Humm-Wadsworth unfavorable, while the Bernreuter receives both praise and disapproval.

Scattered comments on other points of interest about personality questionnaires are these:

They are difficult to improve beyond the present level. It will take a first-rate genius to make any great improvement. A lot of effort has been expended lately, with little progress resulting.

Moderately satisfactory at extremes of distribution. Principal value impresses me in terms of serious deviation from norm rather than as absolute score values.

The psychologically sophisticated person who has some motive for making a good impression can consciously distort results; disturbed person doesn't know the true answers with respect to himself.

Such tests should always be supplemented by other data, such as observations, anecdotal records, and projective techniques. Personality inventories are often excellent aids in "screening" individuals for further detailed observation and study.

Comments Regarding the Rorschach Test

The quantitative replies to Question 2b were tabulated above. The ratings are not greatly different from those pertaining to personality questionnaires save that there is an increase in the number of "Highly Unsatisfactory" ratings in contrast with the "Rather Unsatisfactory," and there is a decided increase in the number of "Don't Know" responses.

The comments with respect to the Rorschach test are notably more vigorous. There are numerous references to "cultism" and "overselling," and even more frequent specific criticisms concerning the lack of validation. On the other hand, a considerable number of the psychologists believe that the Rorschach has value if used clinically by adequately trained persons. Since the Rorschach technique is so much in dispute, it is worth while reproducing a considerable number of evaluative comments. They are grouped below into the two broad divisions just mentioned, plus a miscellaneous set of comments. The parentheses after each quotation contain the rating assigned the test and also indicate whether the respondent considers himself a clinical psychologist or not.

A. Rorschach Test of More or Less Value Used Clinically by Trained Persons

Dangerous for amateurs. A valuable instrument in the hands of a psychiatrist adequately trained in its use. (No rating; clinical.)

I feel this is of value as a strictly clinical instrument in the same way that free association is, but any attempt to objectify scoring of it appears to lead to invalid results. (Doubtfully satisfactory; clinical.)

The Rorschach has already demonstrated its value in clinical usage. The increasing research by "courageous heretics" on modified Rorschach techniques may be expected to produce instruments of considerably more merit than are yes-and-no inventories (Moderately satisfactory; clinical.)

In the hands of a few well-trained experts the Rorschach test may be "moderately satisfactory," but it requires too much highly specialized training under the right supervision to develop the reliability needed in practical work. (Doubtfully satisfactory; clinical.)

If one sticks to the few basic categories, the Rorschach is valuable. I tend to doubt the validity of the detailed analyses

some Rorschach experts are able to make. (Moderately satisfactory; not clinical.)

B. Lack of Adequate Validation; "Cult," etc.

I think more systematic research and less cultism could produce something of value in this particular projective technique, as would be true of any projective technique. (No rating; clinical.)

Do not feel "expert" on this item. This test smacks too much of mystery and "cultism." Also rather too esoteric in ramifications for sound scientific appraisal. (Moderately satisfactory; clinical.)

There is need for an empirical validation of this technique. I am impressed by the extent to which its validity is assumed a priori in terms of some semantic scheme. (Doubtfully satisfactory, clinical.)

Highly promising but needs much further research before conclusions can be warranted. Much of present work is scientifically unsound; but some good leads have appeared. (No rating; clinical.)

Too subjective; clinical signs employed are shifted from study to study. Still in the experimental stage and should not be used for practical purposes as yet. (Rather unsatisfactory; clinical.)

As a diagnostic instrument its value is entirely unproved and the Rorschach workers are going about its validation the wrong way: Too much cultism and intuition and too few cold facts! (Highly unsatisfactory; not clinical.)

There has been grossly inadequate validation of the claims for the Rorschach. (Doubtfully satisfactory; not clinical.)

So time-consuming and subjective as unlikely to contribute much that a skillful interviewer would not obtain more promptly by direct means. (Highly unsatisfactory; not clinical.)

Found utterly useless for predicting success in training of aviation pilots. (Highly unsatisfactory; not clinical.)

Those who use the Rorschach seem always to fall under the spell of the special language they have developed and to be more interested in assigning names than in making any extensive and critical investigation of the validity and reliability of their basic concepts. (Doubtfully satisfactory; not clinical.)

C. Other Comments

When, as in some hands, the Rorschach test proves useful I attribute it more to the good sense of the user than to the instrument. (Rather unsatisfactory; clinical.)

This test still leaves much to be desired but is certainly a step in the right direction. If it could be more easily scored and more objective it would be the most effective instrument I know of for clinical measurement. (Moderately satisfactory; clinical.)

From the standpoint of the "buyer" it is worth about 15% of the time one can spend on an individual examination. For the work I do, would practically always want it or a substantial equivalent, for dynamic purposes its potentialities are less than those of T.A.T. (Moderately satisfactory; clinical.)

The Rorschach tests is highly promising. Group administration techniques should make it more widely applicable. (Moderately satisfactory, not clinical.)

A good idea poorly carried out. (Highly unsatisfactory; not clinical.)

Question 3

What do you consider the most promising mental test developments for research students to devote themselves to during the years after the war?

Most of the replies fall into a few broad categories, under which responses are classified below. (In considering the numbers of replies in different classes it should be noted that many respondents listed several ideas; hence the total number of suggestions far exceeds the number of persons answering.)

Most frequently mentioned are needed developments of *new tests*, especially tests of emotion and personality traits. Thirty-six of the 79 psychologists point to work on new tests; 27 of these indicate tests in the field of personality. This result may have been influenced in some degree by the fact that the immediately preceding questions pertained to personality blanks and Rorschach tests.

Illustrative Suggestions Regarding Development of New Tests

Independent measures of ability; specific aptitude tests.

Tests of mental development that evaluate objectively the higher mental processes (as in the Eight-Year Study of the Thirty Schools Experiment of the P.E.A.).

Better non-language performance tests for children and adults.
Better materials, individual and group, for exploring aptitudes of gifted children.

Mental tests for adults, especially older persons.

Short tests for industrial use.

Culture-free tests of general ability.

Tests for special groups—bilingual, blind, deaf, etc.

Wisdom, thinking, judgment, etc., as distinct from mental alertness.

Development of tests for such traits as perseverance, ability to supervise (leadership), emotional and social maturity.

Objective personality tests; indirect scoring so that subject doesn't know its real purpose.

Projective techniques as personality tests; objectification of projective tests

Personality tests using operationally defined concepts tied to particular fields and giving up the magic of such alleged traits as "extroversion," "dominance," and the like.

Measurement of personality factors and of the as-yet-unmeasured but important intellectual factors.

One important area in business is to measure "drive" in prospective executives.

Tests for specific personality traits for which good criteria are available.

Measurement of "basic human drives."

Useful and readily scorable interest inventories.

Interest measures through a wider range of occupational, educational, and avocational activities.

Achievement tests through a wider range of life and job situations.

Tests of educational development that yield sub-scores indicating specific aspects of intellectual development.

Trade and proficiency tests geared to specific occupations.

Tests to measure achievement that will actually be functional in normal life, e.g., homemaking, consumer science, health, marriage, child development, social attitudes, labor relations, propaganda analysis, etc.

Mentioned next most often (by 18 respondents) is the need for work on criteria and the carrying on of validation studies to ascertain the relation of particular tests to criteria.

Illustrative Suggestions Regarding Validation Studies and Criteria

More concentration on validation on large representative samplings against adequate criteria, especially with adults against vocational success.

The predictive values of specific tests for specific performances in practical tasks. From these specifics it will be possible to develop data regarding the "types" of tests that predict success in "families" of occupations or activities.

Factor analysis of abilities and of criteria and establishment of satisfactory inter-relationships among the tests and criterion factors.

Decrease emphasis upon new test construction; increase emphasis upon the development of adequate criteria and validation and cross-validation of tests of special aptitude and other predictors against such criteria with samplings of acceptable size.

Development and evaluation of realistic *criteria* for: (a) occupational successes in given job titles, (b) occupational success as general occupational adjustment, and (c) general social and personal adjustment. Of the two needs (work on tests and on criteria) it is considered that the criterion side should be given most emphasis.

By all means investigators should work as hard on good methods of evaluating proficiency on the job as they do on tests. As an end in itself, this has most salutary effects, but it is essential if one is to validate selective tests adequately.

The one remaining category into which many suggestions fall (17 responses) pertains to studies aimed at analyzing and interrelating the component factors of ability—either by factorial methods or otherwise.

Illustrative Suggestions Regarding Factorial Studies

Application of factor analysis results to the construction of differential aptitude batteries, followed by standardization of such batteries on a wide range of schools and occupations.

Identification and measurement of separate factors. Construction of tests which will measure such factors as independently as possible.

Psychological analysis of separate or group factors to supplement or replace the mathematical analysis now so much emphasized. In recent years we have had an orgy of statistical analyses.

Isolation of meaningful complexes or factors.

Refined experimental work on the isolation of mental abilities.

Fundamental research on the best reference variables of intellectual and temperamental aspects of personality is badly needed.

In addition to the above types of reply, the question elicited a variety of other individual answers. The content of a number of these may be suggested by a mere listing of topics:

- Research on the interview.
- Attention to traits which are not measurable.
- Analysis of job requirements.
- Attitude inventories.
- Relationships between successive annual increases of intelligence.
- Changes of intelligence with age.
- Construction of prediction tables.
- Reliability studies.
- More adequate adult norms; greater attention to representativeness of samples.

A few further points of some interest which have not been covered in the above categories are presented in the following quotations:

The development of a general battery that will measure all of the occupationally significant factors and which can be secured for groups of occupations covering the entire occupational range. Such a testing technique is needed for occupational counselling. The findings with respect to groups of occupations requiring similar abilities would also have an important bearing on curricular development. Accomplishment of this task would demand cooperation among various research groups.

Developments designed to encompass all major aptitudes as opposed to particularity of appraisal, i.e., composite description of the complete person. We need test batteries standardized on the same sample with interrelations and differential validities.

The most urgent need is to try many tests of various kinds on the same people. . . . The establishment of unique human profiles is our most urgent need.

Relation of responses as elicited by inventories and questionnaires to variations in behavior under different environment and changes accompanying education and training.

The use of cumulative records of comparable test data throughout the school and early employment life of the individual, with equal or greater interest on cumulative anecdotal records of actual behavior.

CIVILIAN TESTING IN THE QUARTERMASTER CORPS

W C KVARACEUS AND W. N. DUROST
Civilian Testing Section, ASF

AND

R. F. McCLELLAN
Office of Quartermaster General

THE QUARTERMASTER CORPS is one of the several supply services in the United States Army. The duties of the Quartermaster Corps comprise the initiation, procurement, supply, and maintenance of all articles and equipment needed by every soldier or necessary to the administration of the United States Army with the exception of the weapons with which the soldier fights and certain classes of transport. In order to supply, feed and clothe some nine million men quartered in both hemispheres, the Army maintains 22 Quartermaster and Army Service Forces Depots in the United States and draws heavily on the civilian worker to assist in this vital phase of the war effort. Roughly, 80,000 civilians are employed in some 300 different jobs in these depots throughout the country. These civilian jobs range from the highly skilled technical and professional positions to those of unskilled laborers.

With the advent of the war Quartermaster Corps, like other technical services, expanded to fantastic proportions in the way of a world-wide supply organization. This tremendous growth demanded the hiring of thousands of workers and presented many problems of assignment, training, and employee relations.

Lieutenant General E. B. Gregory, the Quartermaster General, was quick to recognize the basic management principle that results are achieved through people. Since civilian employees comprise a large percentage of the total personnel in the installations under the jurisdiction of the Quartermaster General,

the accomplishment of the mission of the Quartermaster Corps depended, to a very large extent, upon the effective management and utilization of civilian personnel. As one means toward this goal, the Quartermaster General, through Colonel Eugene G. Mathews, Chief of the Civilian Personnel Branch of the Office of the Quartermaster General, in close cooperation with the Civilian Personnel Research Sub-Section, Personnel Research Section, Classification and Replacement Branch, Office of the Adjutant General, has encouraged the proper use of psychological tests. It was felt that these tests could provide concrete evidence concerning employee knowledge, aptitudes and skills for use particularly as an aid to Placement, Training and Employee Relations activities. As a result of much thinking and considerable experimentation in a number of depots, a Civilian Testing Section has been set up within the Civilian Personnel Branch, Personnel Division at OQMG to:

- a. Encourage the use of ability, skills, and aptitude testing in all Quartermaster and Army Service Forces depots employing civilians, and to advise in the establishment of a Testing Section as a component part of the personnel organization.
- b. Coordinate and standardize all testing activities currently being conducted and proposed in all Quartermaster and Army Service Forces Depots.
- c. Render technical and staff assistance to all Quartermaster and Army Service Forces Depots through the issuance of a testing manual that will serve as the official guide in the use of tests and testing materials and specific Quartermaster testing policy and procedure.
- f. Compile for the Quartermaster General such progress reports on testing as may be requested.

Cooperation of Office of The Adjutant General

At the request of Colonel Eugene G. Mathews, Chief of the Civilian Personnel Branch, Office of The Quartermaster General, two technicians¹ were assigned to the Quartermaster

¹ Dr W. C. Kyvaraceus and Dr W. N. Durost, who were attached at this time to the Field Staff, Civilian Personnel Research Sub-Section, AGO.

Corps to assist in planning and setting up testing programs in selected depots and at the headquarters level. This was to provide a background of experience on the basis of which to set up service-wide testing, if the experimental program showed any worth-while promise in solving some of the personnel problems facing the Placement, Training and Employee Relations officials. The office of the Adjutant General already had constructed tests of general intelligence, clerical and mechanical aptitude, and knowledge and skills and had been given clearance for use of all testing materials prepared by the United States Employment Service. Machinery had also been set up within the Civilian Personnel Research Sub-Section of the Adjutant General's Office to construct additional tests whenever the need for such tests was shown. In general, the development of a systematic and comprehensive service-wide testing program throughout the Quartermaster Corps has been done with the active assistance and cooperation of the Office of the Adjutant General.

Trial Testing in Selected Depots

Some testing already was going on in several depots² before technical assistance was procured from the Office of the Adjutant General. This testing was usually an adjunct of either the placement activities, the training program, or the employee relations activities. Often, the testing was spotty and haphazard, and seldom was a qualified full-time or even part-time technician in charge. The testing activities in these depots, however, revealed an awareness of the fact that some assistance could be obtained in the personnel program through the wise use of psychological tests.

Personnel technicians from AGO visited two Quartermaster Depots to set up testing activities. In each instance, qualified technicians with adequate professional and clerical staff were recruited to head the program in the local installation. In another depot, the testing activities already had been set up under the Placement Branch. In one of the new installations

² Credit is due the Philadelphia Quartermaster Depot which, under its own initiative, had activated a comprehensive testing program prior to headquarters planning

Testing was placed under Training. In the other, Testing was established as a separate unit coordinate with Placement, Training, and Employee Relations. The head of this separate testing unit was made immediately responsible to the Civilian Personnel Officer. The latter type of organization was finally recommended and adopted as most promising if a depot-wide program were to function with the maximum effectiveness.

As soon as the staff was recruited in a depot, attention was turned to the application of testing as an aid in the solution of operating problems. In one of the initial depots it was found that important reassignments and promotions were to be made within the Shoe Inspection Section. The local testing technician, aided by the AGO representative, prepared a battery of tests which was given to all shoe inspectors attached to the depot. This battery included a learning ability test, a shoe inspection information test which was constructed for the purpose and published by AGO, a man-to-man rating scale, and an activity preference questionnaire, also specially constructed. After the tests were given and the data summarized for the operating officials, further assistance was rendered in utilizing the test results in specific personnel actions. For example, the five best all-round men were selected from which one was later chosen for a special assignment.

In another depot, file clerks were tested with appropriate instruments to discover those clerks whose filing skill was low and who were largely responsible for "messy filing conditions." Those file clerks who showed limited alphabetizing skill, but who did reveal high learnability, were assigned for training; the file clerks who lacked aptitude for this job were re-tested with other clerical batteries and were reassigned to jobs for which they showed more promise. In still another depot, where considerable difficulty had been experienced in the Fiscal Branch due to numerous errors in arithmetic processes, all fiscal clerks were given arithmetic tests to discover the individuals who might be largely responsible for the recurring arithmetic errors. Again, according to the test findings, clerks either were retrained or reassigned. At the same time, testing of all incoming employees was started immediately as an aid

in the assignment of personnel to specific jobs at grade as determined by the Civil Service Commission.

The keynote of the field service was the development of a testing program which would serve as an aid to solving depot problems and, at the same time, demonstrate the potential value of test results in guiding personnel action. These trial testing programs rapidly expanded and became an integral part of the depot organizations. With this experience and the conviction that the wise use of tests could materially aid these and other depots, the establishment of a Testing Section at the headquarters level, Office of the Quartermaster General, was accomplished. The purpose of the newly established Testing Section was to coordinate, encourage, and advise in the establishment and functioning of testing sections throughout depots under the jurisdiction of the Quartermaster General.

The Contribution of Testing to the Total Personnel Program

On the basis of the experience in these depots, testing was conceived as a service (staff) function standing in relation to the Civilian Personnel Officer in much the same way that Depot Control stands to the Commanding Officer. Tests could be helpful in that they provide objective evidence in the form of test scores upon which personnel action could be based. Some of the personnel actions to which testing was found to make a notable contribution were as follows:

Placement of incoming personnel. Although the Civil Service Commission reserves the right to certify employees as to grade, it does not attempt to specify to what specific duties within a given grade an individual is to be assigned. This leaves the local depot considerable latitude in placing new persons on jobs for which they are best suited. The local depot itself must test new people if test scores are to be made available for most effective placement purposes. A variety of tests may be used for this purpose, but basically most Quartermaster Depots found that a limited battery of aptitude and achievement tests could serve most purposes satisfactorily.

Reassignment of personnel at grade. Reclassification of personnel is a very serious business, regardless of the direction

of the change, whether it be promotion or demotion. Testing can do much to support such personnel action by demonstrating the presence or the absence of the desired skills, knowledge, or abilities.

Selection of personnel for training. Far too much of the training being carried on in the operating installations was found to be haphazard in the sense that a general order was issued to train persons in some specific area such as the use of the War Department Shipping Document or military correspondence, without knowledge of which of the persons selected for training already knew the material to be covered by the course. In the case of those whose knowledge was incomplete, there was no evidence as to the specific areas where gaps in knowledge existed, so that the subsequent training was necessarily on a general rather than a selective basis. By the judicious use of specific information tests in such situations, three things were accomplished. (1) Those with a mastery of the information sufficient for the needs of their job were excused from training. (2) The training of those lacking such basic knowledge was justified to their supervisors on the basis of objective evidence. (3) The training was directed to meet the areas of the greatest need. Following training, retests revealed the extent to which training had been successful in imparting basic information. Note should be made here that the failure to get information across to a group may be due to a variety of causes. Testing alone will *not* reveal the reason for failure, but only its existence.

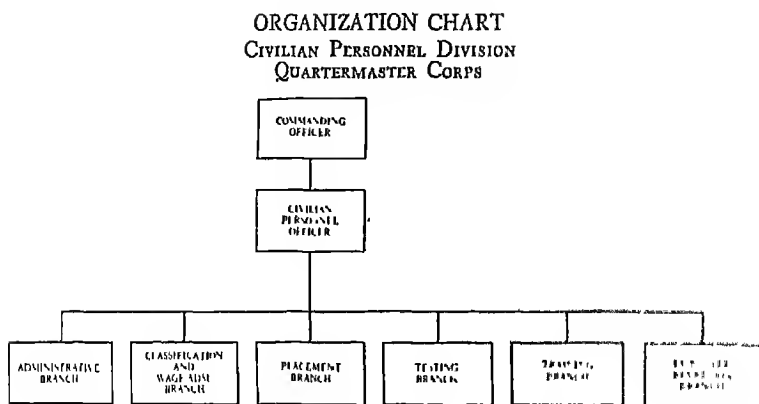
Substantiation of claims of supplementary or higher skill. Many manpower utilization programs listed employees' supplementary or secondary skills. This information generally came from an interview with the employee. Unless the information thus obtained was substantiated by the use of objective tests, little dependence could be put upon individual claims in selecting persons for higher level jobs.

6. *Separation for cause.* When it was necessary to remove an employee for cause, such action was greatly strengthened by the use of objective tests of an informational or work-skills type if the cause was inability to do the work.

The Place of Testing Services in the Depot Organization

Since testing is a staff function, serving all branches within the total personnel program, it cannot fully perform its functions if it is tied to any one of these branches. This has been conclusively demonstrated in all the experience of the Quartermaster and Army Service Forces Depots. When Testing was tied to Placement, its energies were largely absorbed and its policies determined for the most part by the needs and problems of the Placement Branch to the exclusion of Training, Employee Relations, and Operations. When Testing was tied to Training, it was found to be similarly handicapped by having the focus of attention placed on the needs of the Training Branch to the exclusion of the needs of the other branches. When Testing was tied to Employee Relations, it was inclined to take on a guidance aspect, which was not always in the best interests of Placement and Training. Only when the Testing Unit was independent and autonomous, its chief reporting directly to the Civilian Personnel Officer, could it avoid suffering from the restrictions in its activities that were invariably associated, in the minds of operating personnel, with the specialized activities of the branch to which it had been tied.

The desired independence for the testing functions, it was found, could be secured in several ways, two of which were recommended. First, Testing could be set up as a separate branch coordinate with Training, Placement, Employee Relations, and Classification. Second, Testing could be set up as an adjunct of the office of the Chief of Civilian Personnel, in much the same way that Depot Control reports directly to the Commanding Officer. The first plan was approved and forms the basic pattern in most Quartermaster and Army Service Forces Depots. It was recognized that the particular pattern of organization arrived at for any given installation should be determined in the final analysis by local factors. But always it is backed up by the recommendation from the top echelon that the testing activity be given the necessary independence to permit it to do its work unhampered by subordinating it to other personnel functions.



Another major reason for this organizational pattern was the fact that the quality of personnel needed to operate an effective testing program is at least on par with, if not superior to, the quality of the personnel required in any other personnel function. It is not good management to subsume one activity under another when the chief of the subordinate activity is classified as high as or higher than the chief of the parent branch.

The Staffing of the Testing Branches

It was soon discovered that the type and number of personnel needed for a testing program depended on the number of civilian employees to be served and the variety and complexity of the jobs filled by the civilian employees, as well as on the strength of the personnel program in operation in the particular installation. Considerable care has been given to the recruitment of trained and experienced personnel to head the testing activities in the depots. As the first step in developing a promising testing branch, a trained personnel technician was hired. Obviously, the successful functioning of a testing program will center around the qualifications and experience of the personnel hired for these strategic positions. The duties and responsibilities of the personnel technician in charge of the Testing Section or Branch were such that the position always was classified at a professional level, and if the depot was very large (5,000 to 8,000 civilian employees)

a fairly high professional grade was called for. These jobs were set up in accordance with the specifications outlined by the Office of the Secretary of War. Usually the personal technician in charge of testing had one or more professional assistants, depending again on the size and type of the depot. One or two clerks rounded out the office force. The following job description, taken from a typical depot, presents a concrete picture of the type of activity involved and the type of personnel recruited.

Personnel Technician (Testing) P-3

Supervision received Works under the general administrative and technical supervision of the Director of Civilian Personnel, but with considerable latitude in planning the details of specific assignments, and with the responsibility of carrying out these assignments without detailed direction as to technique.

Supervision exercised From time to time will supervise varying numbers of clerical and technical personnel engaged in administration, scoring, and item-analyzing of tests, rating scales, questionnaires, etc., and in the compilation of data growing out of the use of such instruments, the computation of necessary statistics, and the preparation of reports. Will be responsible for the training of such clerical personnel in the necessary techniques, when personnel with previous experience are not available.

Duties and responsibilities At the X Quartermaster Depot (a Class IV installation employing several thousand civilians in separate operating units) is responsible for making the preliminary selection of tests, rating scales, questionnaires and other devices for use in meeting specific employee personnel problems. Such preliminary selection will be subject to the approval of the Director of Civilian Personnel. Is responsible for the administration of such tests, rating scales, questionnaires, and other devices to the personnel selected, either administering the instruments personally or training clerical personnel to do such administration. Is responsible for maintaining reasonable working conditions in the space provided for the administration of such instruments, with respect to lighting, ventilation, freedom from interruption, etc. Is responsible for the scoring of these instruments, for the summarization of the data so obtained, for its interpretation to the operating officials who will use the information (Placement, Training, Employee Relations, Classification, Operations), and for the preparation of reports at periodic intervals and on special occasions as required. When a test or other instrument must be selected for measuring aptitude or skill in the performance of the duties of some specific position, the Personnel Technician is required to familiarize himself with the details of the position involved by consulting with the Classification Analyst or by making job analyses, or by consulting the Testing Section of the Office of the Quartermaster General or other agencies such as the Adjutant General's Office. If no test is available, may be required to construct a suitable instrument. When a test or other instrument is required to cover some specific body of information, such as the nature of and the regulations covering the use of the War Department Shipping Document or Procedure, is required to consult such sources as enumerated above to discover the existence of such a test, and if none is available, construct one to fit the need. Is expected to construct tests, rating scales and questionnaires from time to time in connection with the selection of personnel for training and the measurement of achievement after training. In all such test construction work, the procedures used must be acceptable from a professional standpoint in line with recent developments in this field. In the analysis of test data or data obtained by use of questionnaires, rating scales, or other similar means may be required from time to time to compute means, standard deviations, correlation coefficients of various kinds, reliability and validity coefficients, to prepare bar diagrams, percentile curves, histograms, etc., and

to set up local percentile or standard score norms and perform other related duties as assigned.

The importance of developing a working testing program which is firmly set on a concrete foundation of adequate personnel cannot be over-emphasized.

Tests and Test Batteries

All testing for placement purposes and a greater part of all other testing in the depots is done with some particular job in mind. Either the person tested is being considered for a definite opening or the adequacy of his performance in a particular position is being appraised. There are too many jobs in the Quartermaster Corps to permit the establishment of a recommended battery of tests for each job separately. However, it is possible to recommend test batteries for a series of positions, such as certain series in the clerical or mechanical fields. In a few cases specific batteries were set up for job classes or for specific jobs. This has been done for those positions which appear with the greatest frequency in the various depots.

The basic test³ usually given to every employee who is hired, or who is referred for testing for any reason, is the *Learning Ability Test*, which exists in two forms. This is a general verbal abilities test, omnibus type, using multiple-choice items, which closely resembles a general intelligence test such as the *Otis* type. In cases where there is a language handicap or a question of illiteracy, a non-reading intelligence test is substituted. The next most widely used test is a clerical aptitude test, which is given in total or in part to all persons in clerical positions.

Other tests commonly used in the depots include the following: *Number Speed, Typing, Shorthand, Military Correspondence, Digit Reversal, Word Meaning, Coding; Clerical English Battery*, including tests of *Abbreviation, Capitalization, Compound Words, Grammar, Punctuation, Spelling, Word Division in Typing, and Word Selection; Mechanical Aptitude and Technical Aptitude Test*, including the following: *Mechanical*

³ All AGO tests are restricted materials. Commercial tests are used in the Quartermaster Civilian Testing program only when AGO tests are not available.

Knowledge, Visual Discrimination, Space Relations, Inspection Speed, Technical Mathematics, Technical Reading, Figure Cancellation, Elementary Electricity, General Automotive Information, and Radio Information Tests. In addition, specific tests of knowledge on the War Department Shipping Documents and the Vendor's Shipping Document have been prepared. Tests for warehouse jobs, such as packers, checkers, and storekeepers, have also been constructed. Most of these latter tests are used primarily in the training program.

The Division of Occupational Analysis, War Manpower Commission, has authorized the Adjutant General to reprint or adapt, on a restricted basis for Army use, the tests constructed for the United States Employment Service. At the same time, the *Oral Trade Questions* have also been made available through official channels. In all, several scores of tests are available for use.

Attempts are being made to set up batteries of tests for some specific jobs. Norms are being gathered in terms of the performance of new employees and in-service employees. The following batteries are given as examples of specific batteries prepared for specific jobs.

Clerk-Stenographer

- Learning Ability Test
- Clerical Aptitude Test
- Typing and Shorthand Test

Checker

- Learning Ability Test
- Clerical Aptitude Test
- Number Speed Test

Inspector of Clothing

- Learning Ability Test
- Inspection Speed Test
- Optical Precision Stereoscope Test
- Rate of Manipulation Test
- Color Perception Test

Fork Lift Operator

- Learning Ability Test
- Eye, Hand, Foot Coordination Test

Physical Fitness Tests

Vision

Endurance

Hearing

Inspector of Shoes

Learning Ability Test

Inspection Speed Test

Optical Precision Stereoscope Test

Quartermaster Shoe Inspection Test

Contract Negotiators

Learning Ability Test

Clerical Aptitude Test

Critical Thinking Test

Arithmetic Reasoning Test

Personality Questionnaire

Baler

Learning Ability Test

Revised Army Beta Test

Rate of Manipulation Test

Physical Fitness Test

These are examples of specific test batteries assembled in terms of the actual skills involved on the job. Some of these batteries are now in the process of validation.

Test Records and Reports

It cannot be emphasized too strongly that *testing is a service function that has no value unless the tests are used*. Hence, the system of test records and the method of interpretation is aimed at the purpose of maximum utilization of test results.

Raw scores are never given to operating personnel or to anyone outside of the Testing Branch. Various types of norms are available, and their use depends upon the purpose to which the test scores are to be put. The most generally used type is one in which the total group upon which the norms are based is divided along the base of the normal curve according to a five-point scale. Each of these groups is rather easily characterized in terms of adjective phrases, indicating degrees of

goodness in the quality, skill, or ability measured by each test. A test report with these descriptive phases is made out for each person tested, and usually turned over to the Placement Technician, Employee Counselor, or Training Director. In other words, the usual procedure is to give the operating personnel an interpretive comment on the test results rather than the test scores themselves. More detailed norms are available in the Testing Branch for use in cases which call for closer interpretation.

The test results (raw scores) are recorded on a Test Record Card which is filed in the Testing Branch. A key-sort type of card is the recommended record card used in most Quartermaster and Army Service Forces Depots. At the same time, test results in terms of five-step grades are entered upon the Employee's Qualification Card which is maintained by the Placement Branch. This card is always consulted whenever any personnel action is contemplated. Considerable use is also made of percentiles as a further interpretive score.

A daily log of the test results of all individuals examined is recorded in duplicate. One copy of the daily results is forwarded to Headquarters monthly, where the results are studied and service-wide norms are developed. The local installation uses its copies of the daily log to establish local norms.

Validation of Tests and Establishing Critical Scores

Attention has been given to the validation of test results and to the determination of critical scores. Various types of criterion data, such as rating scales, quality of work output in terms of error scores per unit of work, and quantity of work output, have been employed in these studies. Some of the investigations have been discouraging in their results, especially in the use of rating scales. Efforts are now being made to use various types of criterion data other than rating scales, to show the relationship between test scores and job performance. It is felt by the writers that the difficulty in obtaining satisfactory validation data and satisfactory critical scores in many cases has been due much more to the inadequacy of the criterion data than to the selected tests.

Quartermaster Testing Handbook

In view of the establishment of general testing policy of the Quartermaster Corps and the expanding testing programs throughout the installations, a handbook on testing has been prepared. This handbook is divided into two main parts. The first part discusses the role of testing in terms of the contribution that testing can make to various phases of the depot program, including the responsibility of the Commanding Officer, the Civilian Personnel Officer and the Chiefs of Training, Placement, and Employee Relations. The second part discusses the more technical phases and procedures of testing and is intended primarily for the personnel who make up the staff of Testing Sections. The manual describes the Quartermaster Testing Program in considerable detail and reflects the experiences gained in establishing Testing Sections or Branches in several installations.

Summary

The service-wide testing program which has been planned and implemented from the headquarters level in the Quartermaster Corps shows considerable promise. It may well point the way in industrial testing, not only for other technical services in the Army Service Forces but to industry as well. The place of testing has been carefully defined as an inherent part of an over-all personnel program on a service-wide basis, involving some 80,000 civilians employed in hundreds of different jobs.

Briefly stated, the functions of the Testing Units as set up within the installations under the jurisdiction of the Quartermaster General are as follows:

1. Select appropriate batteries in terms of job requirements.
2. Administer, score and interpret all tests used in connection with placement of new employees, determination of training needs, evaluation of training, and employee relations activities.
3. Conduct testing surveys at the request of operating officials, and provide suitable reports of results through proper channels.

4. Conduct the necessary research to establish local norms and to determine the validity of experimental tests.
5. Construct new tests as required.
6. Maintain necessary records.
7. Coordinate testing with other personnel activities.

Testing is only one aspect of the Civilian Personnel program in the Quartermaster Corps. But it is an important aspect. It provides valuable information concerning the employee's abilities, aptitudes and skills, obtained in a relatively short period of time. Tests, wisely used and with due consideration to their limitations, enable the immediate location of the more apt workers and the more trainable employees. Tests are making an important and vital contribution to the war effort by insuring the maximum utilization of manpower. It can truly be said "Tests have gone to war on the civilian as well as the military front."

TESTING BY MEANS OF FILM SLIDES WITH SYNCHRONIZED RECORDED SOUND

HERBERT A. THELEN
The University of Chicago

1. *Preliminary Considerations*

FUNDAMENTALLY the method of evaluation is to put the student into situations likely to result in experiences engendering overt responses which can be used for valid prediction of behaviors assumed to constitute the goals of education. The implementation of this method is seen to require a clear statement of educational objectives, the setting up of controlled patterns of stimuli appropriate to the level of maturity and other characteristics of the students, the description of overt responses of each student, and the generalization of the descriptions of these responses into predictions of types of response in a wide range of similar situations, and finally, the evaluation of the degree of appropriateness of the responses of each student as compared with those of his classmates.

For the purposes of this discussion of a new type of test, we shall postulate that by a "test" we mean an instrument which measures status of a student relative to certain objectives and relative to a group of students tested at the same time. We shall further assume that the instrument is constructed after the specific objectives have been described operationally, and that the test situations have been formulated to elicit behaviors to be appraised with due regard for appropriateness of subject-matter content, level of maturity, and type of problem tension engendering the observed responses. These assumptions are required in order that different media of tests may be compared.

A test is of little value if the results of testing cannot be interpreted for some clearly stated purpose. Probably the

major factors in the interpretation of a test to which the above postulates apply are: (1) adequacy of available description of the testing situations, (2) degree of insight of the interpreter into the psychology of learning and behavior, and (3) knowledge of the culture, maturity, tool skills, familiarity with situations similar to those in the test, and other relevant attributes of the students tested; for these three factors determine the validity¹ of description of the behaviors presumed to account for the responses checked in each test item. Certain aspects of these factors may now be considered in detail.

A. The Testing Situation. At the operational level, a test is generally a piece of paper with writing on it. For most students it presents the dominant stimuli in a total situation which includes the test administrator, a group of students, the physical environment, and the temporal place of this situation with respect to an undefined sequence of experienced situations, both prior and subsequent, each of which has engendered or anticipated a more or less relevant experience of the student. The only factors which may be clearly described in the testing situation relate to the test itself, and, with lesser comprehension, to the procedure for administering the test. If the test actually does elicit the full attention of the student, then the remaining physical factors are assumed to be of negligible importance in determining the score, and the factors of relevant previous experience are assumed to be the major cause of differences among the scores of the students.

It follows that to describe the testing situation adequately, one must know: the effectiveness of motivation of each student, the exact procedure for giving instructions (and making sure that they are understood), the relevant physical and social conditions during administration of the test, and, finally, the nature of the test items (language used, problem-tensions aroused, attitudes implied, slogans employed, and the like). Since the extent of the motivation of a student cannot be measured directly, and can be inferred only in the case of some tests giving a pattern of scores, it is customary to assume

¹ Defined as amount of correspondence between predicted and observed behaviors in the entire population of problems sampled by the test.

that all students are either equally or maximally motivated. Since the extent of a student's understanding of the directions for taking the test is not measured apart from his achievement on following the directions, it is assumed either that the understanding of directions is perfect or else that ability to understand directions is part of the achievement being appraised; the latter is probably the sounder assumption. The usual assumption about physical and social conditions is that good lighting, adequate ventilation, and an empty seat between adjacent students are about the only relevant factors. Since a test should not be given if it is inappropriate for the group being tested or if it does not measure the desired objectives, it follows that use of a test assumes that it will be valid under the conditions in which it is used. This assumption can be justified only by detailed analysis of the individual items with reference to such factors as those suggested above. The necessity for making these or similar assumptions confronts the interpreter of any test. The justification for these assumptions may, however, vary from one test situation to another. It would be desirable to make these assumptions as valid as possible.

B. "Verbal" versus "Real" Situations. If a test is used merely as a hurdle to be cleared for the sake of a certificate, then, within wide limits, the nature of its items makes little difference so long as the distribution of scores covers a wide range and the students opined to be "best" receive the highest scores, and the students believed to be "poorest" receive the lowest scores. Such a test used in this way does not require discussion here because the preliminary postulates do not apply to it. The statement that "specific objectives have been described operationally" means that the major, generally-stated objectives have been analyzed and broken down into a large number of distinguishable types of action. Each of these types has its own properties and relationships with other types; these relationships constitute the rationalization of the general objective. In a course for which such an instrument is appropriate as an aid to evaluation, there has been some effort to teach the students the types of behavior, the criteria by which the

most appropriate type may be selected in a given situation, the rationalization of the general objective, and the scope of applicability (range of situations and purposes) of the objective. These objectives are generally understood to refer to a wide range of everyday experience; but their measurement is essayed through symbolic verbal experience. Under these conditions it is not difficult to come to the conclusion that all mental processes come under the heading of "reading comprehension" and that therefore the major task of schools is to teach students to read.

Situations as conveyed by paper-and-pencil test differ from the situations encountered in everyday life in some very important respects, and these also present the test interpreter with the need for making a number of assumptions whose validity is generally not easily appraised. Some of the factors involved are:

(1) Ambiguity. Through verbal symbols standing for objects one attempts to engender the same response that the objects themselves would produce. Since object-names stand for classes rather than for individuals, a number of qualifying adjectives (or stated properties) must be given to specify the individual. But the adjectives themselves are usually associated with outstanding types of objects to which the adjectives best apply (within the experience of the reader), so that the reader is confronted with the task of visualizing a specific pattern of aspects (which in turn can never completely reproduce an object) from what amounts to a series of generalizations about varieties of types of objects. It would be quite surprising if such a description meant the same thing to two different students!

(2) Semantic misrepresentation. The particular pattern of aspects through which the tester attempts to convey the object to the reader's experience may or may not coincide with the pattern of aspects by which the reader symbolizes or would symbolize the object.

(3) Incompleteness of pattern of stimuli. All the senses together enter into the experiencing of an object or situation. Presumably the number of possible kinds of response would

depend partly upon the complexity of the pattern of stimuli (in this case represented symbolically). There is no adequate symbolic notation to even represent odors, tastes, or surfaces; and therefore aspects ordinarily obtained through these types of perception cannot be given. This is one kind of cultural limitation imposed upon test items. It may be argued that this limitation is unimportant because we make little use of perceptions from these senses (since they cannot be represented—a vicious circle), yet there is undoubted depreciation of the validity of the vicarious experience engendered within these limitations.

(4) Non-simultaneous presentation of the pattern of aspects. In reading, images are built up piecemeal, and with extremely attenuated impact. Furthermore, the order of presentation of the components from which the pattern is formed is fixed. All three of these factors are artificial and may be expected to give the test situation different meaning from that of the corresponding “real” situation.

(5) Selection of aspects. Any symbolic system of representation proceeds by selection of relevant aspects. This focuses or fails to focus attention upon details whose importance is thus magnified or diminished out of proportion to the other details in the situation.

The above theory is presented to rationalize by means of symbolic representation some of the experienced difficulties with paper-and-pencil tests. In general, it may be said that these tests present artificial situations to which the range of kinds of response is limited, and that facility in manipulation of verbal symbols is an important factor which masks to some unknown degree the nonreading abilities to be measured. The use of such tests has led to emphasis upon learning of self-contained relationships among symbols rather than upon phenomenal aspects represented by the symbols—students are taught “maps” rather than “territories.”

It seems reasonable to suppose that instruments dealing with experience solely at the verbal symbolic level may, nevertheless, be of some use in evaluating abilities which are defined in terms of behaviors guided largely by verbal maxims or

conventions. Although the situations eliciting the behavior are subject to the difficulties outlined above, the behaviors in these situations may be largely explained as manipulations with verbal criteria.

The behaviors encompassed under the generic title of "critical thinking," because of their high symbolic loading, can be appraised much more satisfactorily than can "attitudes." Thus judgments in terms of stated or unstated criteria, or in terms of logical rules or scientific methodology may be elicited with paper-and-pencil items. The major discrepancy may here be found in that the behavior starts after verbalization with the verbal items, but before verbalization with the non-verbal items. The appraisal is then limited to a part of a process rather than to a whole process. If this limitation is recognized, however, interpretation may be apparently sound.

One obvious way to overcome the difficulties inherent in a verbal presentation of situations is to place the student in a more or less controlled "real" situation and then observe his behavior. This sort of technique usually involves individual testing of each student by a specially trained observer; the results may be expected to be less reliable but more valid. Even here, however, the testing situation differs from the population of situations about which we wish to make predictions in that the problems are formulated or at least suggested by the observer; that is, the tension resulting in problem-solving behavior of the testee is stimulated by the observer rather than by the configuration of naturally occurring elements within the situation itself. This technique is admittedly cumbersome and expensive.

The present investigator has become interested in the possibilities of the sound-slide medium for reducing the loading of verbal symbolism and increasing the participation of students in testing situations. The test so far constructed will be described and then tentatively evaluated by means of some of the concepts stated above.

2. *Description of the Test*

A. *Nature of the Instrument.* The test to be described simultaneously presents a controlled pattern of stimuli approxi-

mating "real" situations. To this extent it resembles the controlled observation technique more than that of the paper-and-pencil test. The overt responses of the students are limited to recorded judgments, opinions, and analysis, and to this extent the test is akin to a paper-and-pencil test rather than an observational type of evaluation.

The particular instrument to be described is concerned with the area of behaviors generally referred to as "ability to apply principles." The principles are taken from among those ordinarily studied in the fifth-grade physical science units at the University of Chicago Laboratory School.

The test consists of a film strip, a recorded transcription, and answer sheets for the students. The test is given in a semidarkened room; the light is adequate for students to follow the answer sheet and dim enough for the projected pictures to be clearly seen. The film strip is projected one frame at a time, and the pictures are changed at a signal recorded in the transcription. The recording provides some narration for each situation, authentic sound effects, and directions for marking each item on the answer sheet. Sixteen-inch records are used; these are played at 33 1/3 r.p.m. and run for seventeen minutes. The film slide strip presents one to five pictures per problem, and also contains photographed typewritten titles. Possible answers are presented as depicted right and wrong ways of doing certain jobs, written explanations or principles, depicted members of analogies, depicted illustrations of operation of principles, and other devices deemed appropriate to the specific objective being tested. A few of the problems require brief written statements.

The present test was given in Grades V, VII, VIII and X. The operator stopped the transcription to allow sufficient time for all the students to record their answers for all the items. The pauses provided in the transcription were approximately correct for the tenth-grade students, but had to be lengthened for the others. The test required thirty-three minutes in the tenth grade and about forty-eight minutes in the fifth.

B. The Test Items and Objectives. The test presents nineteen problems focused on ten specific stated objectives in

the area of application of principles in elementary science. The following brief description of the objectives and items may suggest sorts of possibilities for testing with this medium:

Objective I. To recognize a practical (unstudied) application of a principle studied in class.

Problem 1: A ruler clamped in a vise is plucked, and the sound is heard. Then a shorter ruler is plucked, and the sound is heard. The students are asked to consider three depicted ways of getting different notes from a violin: by turning a peg, playing open strings, or playing up and down the scale. For each of these three situations they record that the notes are different for the same reason that the notes from the ruler are different, that the notes are different for some reason other than that shown with the ruler, or that there is not enough evidence to decide between the alternatives.²

Objective II. To arrange events in a temporal sequence according to a developmental principle.

Problem 2. Three pictures designated *A*, *B*, and *C* show a man using a lathe in the construction of a gadget, using mechanical drawing instruments in designing the gadget, and having an inspiration for the gadget. The student writes the letters *A*, *B*, and *C*, in the order indicating the sequence as he thinks it really occurred.

Objective III. To recognize (from a studied principle) the best technique for solving a simple problem.

Problem 3: Situation: Grease in a skillet has caught fire. Student selects the better depicted method of putting out the fire. Choices: Clapping a lid on the skillet; running water into the skillet.

Problem 4: Situation: Balancing of large weight and small weights hung from a differential pulley. Choices: Large weight hung from smaller diameter; large weight hung from larger diameter.

Problem 5: Situation: Location of tray for fastest freezing in the cooling unit of a refrigerator. Choices: Lower right-hand corner of freezing unit; middle top shelf of freezing unit.

Problem 6: Situation: Position of head to hear a tuning fork loudest. Choices: One ear directed toward the fork; face turned toward the fork.

² This is an analogy between a studied laboratory situation and an unstudied practical experience. The converse objective, recognition of a laboratory setup which operates on the same principle as a familiar practical process or device, was not tested.

In these problems the student selects the better choice, selects both choices as being satisfactory, or rejects both choices.

Objective IV. To recognize situations illustrating the operation of a stated principle.

Problem 7: Stated principle: A smaller force can overcome a larger force provided it moves farther and faster than the larger force. Situations: Boy lifting handles of a wheelbarrow; jackscrew raising a heavy load; single fixed pulley suspending two equal weights

Problem 8: Stated principle. Sound is produced by the vibrations of objects. Situations: Block of wood hit with hammer; water poured from pitcher to glass; whistle being blown.³

In these problems the student rates each situation as illustrating the principle, not illustrating the principle, or as insufficiently described for a decision to be reached.

Objective V. To identify a simple familiar mechanism or process.

Problem 9: Find the wedge. Situations: Driving a nail; sawing wood; pulling a cart up an inclined plane.

Problem 10: Find the sound being reflected. Situations: Boy shouting "around a corner"; man shouting in a large empty room; hammer in piano striking a string.

In these problems the student rates each situation as depicting the mechanism or process, as not depicting the mechanism or process, or as insufficiently described for a decision to be reached.

Objective VI. To compare predicted (from familiar, unstated principle) results with observed results in simple laboratory situations.

Problem 11: What is wrong with this picture? Situation: China dish is shown being heated by the luminous flame of a Bunsen burner. Then the flame is turned off, and the dish is seen to be clean.

Problem 12: What is wrong with this picture? Situation: Bimetallic bar is shown to be uncurved in a hot flame, and curved after cooling.

³ This problem may also be regarded as a sort of logical tautology, since any case of sound production must "illustrate" the principle. The tenth-grade students were probably sensitive to this aspect, whereas the lower grades distinguished between the whistle (a musical instrument) and the other two sources (noisemakers).

In these problems the student rates the second picture as depicting a correct result, or else explains briefly what is incorrect.

Objective VII. To rate statements of principle or fact as useful in explaining depicted phenomena.

Problem 13: Phenomenon: Whispering is transmitted by a garden hose. Statements: "Some solids transmit sound better than air does." "Sound is reflected by surfaces." "Whispering is higher pitched than talking." "Sound travels outward in all directions through a gas." "Sound is partly absorbed at surfaces."

Problem 14: Phenomenon. Pitch of a tuning fork is the same whether hit hard or softly, whether held in the air or mounted on a resonance box. Statements: "Rate of vibration of an object depends upon its size." "Loudness of a sound depends upon how much the source vibrates." "Sound is produced by vibrating objects." "The number of overtones in a sound depends upon the construction of the source." "The rate of vibration of a string depends upon its tension."

Problem 15: Phenomenon: Man sitting on stepladder in a room is warmer than he would be on the floor. Statements: "The floor conducts heat more rapidly than the ceiling." "Heat rises because it is lighter than cold." "Less dense objects float in a more dense liquid or gas." "A certain weight of air occupies more space when it is hot than when it is cold." "Heat is due to moving molecules and therefore hot things move more rapidly than cold things."⁴

In these problems the statements following the description of the phenomenon are flashed on the screen one at a time. The student rates each statement as being helpful in the explanation, or as not being helpful in the explanation.

Objective VIII. To identify an incorrect postulate in the depicted solution of a problem.

Problem 16: To tune a viola. Depicted solution: The viola is tuned aurally, and the position of each peg marked by means of a scratch. From then on, the instrument is tuned by turning the peg to line up the index and scratch. The experiment is tried, and the viola heard to be out of tune.

Problem 17: To keep cool on a hot day. Depicted solution: A boy is shown putting on successively a sweater, a coat, and

⁴ The criteria or conventions of explanation tested are concerned with relevance, correctness of statement, closeness of analogy, description of mechanism, and the like.

a blanket. In dialogue with the narrator he explains that this will keep the hot air from reaching him, but admits that he is still hot.

In these problems the student writes a brief criticism of the solution shown.

Objective IX. To select the best stated prediction in a practical unstudied situation.

Problem 18: Situation: Taking the nut off a large bolt. A pipe has been slipped over the wrench handle, and a man is pulling on the pipe. Stated predictions "The pipe will bend." "The nut will come off" "The bolt will be twisted in two." "Nothing will happen."

The student indicates the prediction he thinks most tenable.

Objective X. To select the most appropriate opinion (verbal expression of attitude) about the desirability or undesirability of a depicted situation.

Problem 19: Situation: A pile of trash and old newspapers in the corner of a "dark, warm basement" is shown. Four opinions expressing different degrees of alarm over possible danger are given.

The student selects the opinion which he most nearly agrees with.

Problem 3 is depicted in full. The technique of testing includes the following steps.

(1) Presentation of a title or statement designed to gain interest, to indicate the general nature of the task, and to mark the beginning of a new problem. The title is read by the narrator while it is on the screen.

(2) Description and depiction of the problem situation. Pictures are arranged in sequence, and, together with the narration, tell a simple story.

(3) Presentation of answers from which to select. These may be depicted ways of doing things, verbal statements (which may or may not be read by the narrator, depending upon the objective) of explanation or prediction, and the like. In some problems the student is asked to write in his explanation or criticism (short-answer essay).

(4) Giving of directions by the narrator for answering the item.

(5) Allowance of sufficient time for all students to mark the answer sheet.

Steps 2 and 3 may occur simultaneously. Step 4 may precede 3, or even 2.

Film Slides (titles
and pictures)

What is the right way to put out the fire in a skillet?
(See Plate 1)

(See Plate 2)
(See Plate 3)

Problem 3

Recording (narration, directions, sound effects)

NARRATOR: What is the right way to put out the fire in a skillet?

SOUND EFFECT: Bell signal.

NARRATOR: This sort of thing sometimes happens when we heat a skillet with grease in it.

How shall we put out the fire?

SOUND EFFECT: Bell signal.

NARRATOR: Is covering it with a lid a good way to do it?

(Pause)

SOUND EFFECT: Bell signal.

NARRATOR: Or would it be better to pour water on it?

(Pause)

In answer space 5, write *A* if the first way was the right way to put out the fire, write *B* if the second way was the right way to put out the fire, write both *A* and *B* if both were correct, or draw a dash if neither way was correct. (Pause, stopping transcription if necessary, until all of class has marked the answer space.)

C. Results of Testing. By means of this instrument utilizing the sound-slide medium, it was desired to explore a number of possible types of items and situations. The brief sampling of the various objectives precludes expectation of any high degree of reliability with respect to single objectives. Further-

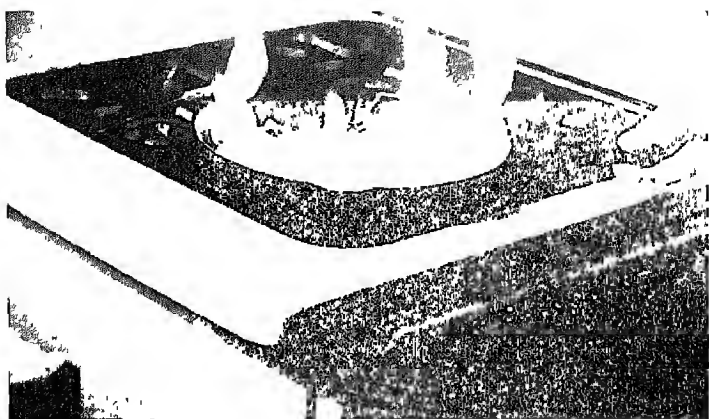


PLATE 1



PLATE 2



PLATE 3

more, if the objectives actually are different types of behaviors, then the test as a whole would not be expected to have high internal reliability. On the other hand, the test could be considered valid if predictions guided by an acceptable theory of learning and based upon knowledge of the learning experiences of the students were borne out by the test results. It was believed reasonable to suppose that this test, presenting the stimuli simultaneously and with a minimum of verbal symbolic representation, should have high face validity. To test this hypothesis a series of predictions about the results of testing were set up and studied. The predictions were:

(1) Accuracy on the test as a whole will increase with the grade level of the students.

(2) Increase in the appropriateness of responses in particular items will spurt between the grade levels above and below which the relevant principles were studied.

(3) There should be some evidence of increasing maturity of thought discernible in the pattern of responses as grade level increases. While such patterns have not yet been described adequately, there should be agreement with such fragments of information as are now available.

The results bearing upon these three predictions are:

(1) Median scores: fifth grade 16.8, seventh grade 19.5, eighth grade 21.0, tenth grade 23.5.⁵

(2) The placement of principles in the science curriculum of the Laboratory School has been relatively constant for several years, but the courses have been taught by several teachers, some of whom are no longer in the school. Knowledge of the learning experiences of the students was consistent with observed spurts in accuracy relative to all the items for which such knowledge was available.⁶ In other words, the test results reflected the learning experiences faithfully so far as they could be described.

⁵ The large influx of students new to the school in the ninth grade may result in a somewhat low median for the tenth grade as compared with the other grades. No attempt was made to match the samples of students because there is no good reason to suppose that the students in Grades V, VII, and VIII are different in any inappropriate dimension.

⁶ About half the responses.

(3) The evidence concerning the changes in pattern of accuracy from grade to grade is meager and subjective. Regarded as empirical finding, it would be worthless, but its consonance with parts of the pattern of anticipations increases the validity of the instrument by some unknown amount.

- (a) The test key calls for two responses that "more information is needed for a decision." In Grade V the accuracy was approximately that expected by chance; there was a decrease in accuracy up through Grade X. This is in agreement with the common observation from Interpretation of Data Tests that "tendency to go beyond the data" increases with grade level (in the absence of special training).
- (b) Accuracy of rejection of the irrelevant reasons in problems 13, 14, and 15 increased markedly between the eighth and tenth grades, but it cannot be shown that subject matter which may have been presented during the ninth grade does not account for the gains.
- (c) The decline in accuracy with principles known to have been studied in the fifth or sixth grades and not reviewed subsequently appeared to depend upon the directness of applicability of the principle as learned.

3. *Criticism and Evaluation of the Medium*

The discussion under "Preliminary Considerations" above provides several criteria which may be used in evaluating and criticizing the type of test here dealt with.

Two facts give us some assurance that the testing situation is controlled and therefore definable to a high degree: (1) All instructions for taking the test are recorded, and this fixes the factor which is usually the most variable in the administration of tests. (2) The test holds the interest of the students and motivates them to work intensively. This is stated as a fact as a result of discussions of the test with the classes taking it, and as a result of observations of behavior of the students while taking it.

The "realness" of the test situations is greater than with paper-and-pencil tests. Consequently it should enable more

valid predictions as to the behavior of students in similar "real" situations, and this type of prediction is assumed to be the most legitimate purpose of achievement testing. The use of motion pictures for depicting some of the situations involving changes along the time dimension would presumably increase the "realness" further (as would also the use of stereoscopic pictures and color. Whether this increase would justify the increased expenditure of effort in making the test is not known; careful analysis of the objectives and situations would enable one to set up hypotheses).

The sound-slide medium very much minimizes the customary use of verbal symbols in conveying the situations; this should make possible the evaluation in the lower grade levels of some behaviors hitherto not readily available for testing. (An illustration is the identification of assumptions in problem 16.) The minimization of reading comprehension as a prime factor in determining the student's responses should also make possible the testing of many objectives more directly.

The more complete presentation of situations by picture and sound means that the pattern of stimuli comes closer to actual experience. Coupled with the advantages listed above may be an increased difficulty of "focusing" items so that the student does not respond unduly to irrelevant stimuli. In other words, the more completely the situation is conveyed, the greater the number of possible types of response, and care must therefore be exercised in stating the question unambiguously so as to elicit the type of response which is most informative in the evaluation of the objective to be appraised.

4. Plans and Suggested Possibilities

Other factors being equal, the more adequately a situation is presented, the more valid the response. It seems reasonable to suppose that this medium may have interesting potentialities for the appraisal of attitudes. Instead of stating an opinion as to preference in verbalized general situations, a student might be asked to criticize a depicted course of action, or to choose among several depicted solutions to a problem involving a conflict in values. Instead of having to select the

relevant aspects of a situation for the student (as one must do in verbal presentation), it would be possible to present subtle but crucial factors disguised in situations. In this event the responses of the student might be governed more by the values he lives by and less by the slogans he has learned.

The science department of the Laboratory School is working as a group on the construction of a sound-slide test to appraise ability to form reasonable conclusions. A variety of situations in and out of science will be used in an effort to find out whether this ability can be described as an entity apart from associated learnings of subject matter. The identification of a number of such abilities plus the development of adequate means of appraisal would make possible some significant research on teaching methods, and might well lead to a complete reorganization of the content of elementary science courses.

5. *Summary*

A new type of test making use of pictures with synchronized narrative, sound effects, and instructions is described. The use of such a test for appraising some aspects of ability to apply elementary principles in science is explored.

Advantages claimed for the sound-slide test are: (1) uniformity of administration of the test from group to group, (2) high motivation of the students, (3) minimization of the verbal element with increased validity of testing some objectives, (4) possibility of appraisal of some fairly sophisticated objectives at low-grade levels.

ANALYSIS OF THE TERMAN-McNEMAR TESTS OF MENTAL ABILITY

F. T. TYLER

The University of British Columbia, Vancouver, B. C.

The *Terman Group Test of Mental Ability* was probably one of the most commonly-used group intelligence tests over the period 1921-1941 (8, p. 33). It is likely, therefore, that the revision of this test will be of considerable interest to school officials. Analysis of the revised form by Terman and McNemar should give valuable information to supplement the manual of directions.¹ The purpose of this paper is to present the results of such an analysis.

The Subjects

The subjects were students in the junior high school at Nelson, British Columbia, where it is the practice to administer group intelligence tests in grades 7 and 9. Approximately 100 students took Form D of the TMcN² tests in September, 1942; forty-nine of these had previously taken the KA tests in 1940 in grade 7. The TG test was administered to 71 of the grade 9 pupils in October, 1942. Form C of the TMcN test was given in February, 1943, to 88 of the grade 9 students for whom Form D scores were already available. Comparisons between I.Q.'s on the various tests are shown in Table 1.

The average TG I.Q. in grade 8 in the Vancouver, B. C., school system was 106 in 1940 (12, p. 106), rising to 115 in grade 12. It is likely, therefore, that the average I.Q. in grade 9 is about 108 or 109. The subjects used in the present

¹ "Careful studies of validity and reliability coefficients and norms presented by test authors are all too rare" (9, p. 16).

² The following abbreviations are used throughout: TG—Terman Group; KA—Kuhlmann-Anderson; TMcN—Terman-McNemar, DIQ—deviation I.Q. and RIQ—ratio I.Q. computed from TMcN tests

study appear to constitute a typical sample, although possibly slightly above average. It should be noted from the table that the average DIQ of 88 cases is three or four points below the average for 49 cases. It seems reasonable to expect, therefore, that the average KA and TG I.Q.'s for the whole 88 cases would be somewhat lower than those for only 49 students, so that the average TG I.Q. approximates that found in the Vancouver schools.

Despite the apparent differences between RIQ's and DIQ's, the correlations between them are very high, being .92 and .94 for Forms C and D, respectively, practically identical with the relationship given in the manual of directions, namely, .92 for Form C. As the authors state: "From these data it will be

TABLE 1
Means and Standard Deviations of I.Q.'s on Various Tests

	KA	TG	Form D		Form D		Form C	
			RIQ	DIQ	RIQ	DIQ	RIQ	DIQ
Date	1940	1942	1942	1942	1942	1942	1943	1943
N	49	49	49	49	88	88	88	88
M	109	113	122	113	115	109	116	110
σ	11.30	10.65	20.62	11.65	19.73	13.40	18.02	13.08
σ	Manual of directions						29.10	17.10

seen that the rank order of deviation and ratio I.Q.'s is very nearly identical, but that the magnitude of the I.Q.'s will vary in increasing amount as one moves away from the mean" (11, p. 10). This also accounts for the fact that the mean RIQ of the present sample is larger than the mean DIQ. The difference in value between a student's two I.Q.'s need not concern the teacher if he understands that a difference is to be expected because of the differences in the standard deviations of the two types of scores. The manual of directions might have been more explicit on this point for the benefit of those teachers who are relatively unfamiliar with the meaning of a standard deviation. The authors recommend the use of the DIQ, but the RIQ will be used in many schools because teachers are more familiar with its definition.

*Findings*1. *Correlations between Various I.Q.'s.*

Table 2 shows the correlations between the various I.Q.'s obtained in the present study.

TABLE 2
Correlations between I.Q.'s

Variables	Interval (in years)	N	r*
KA—TG	2	55	.79
KA—DIQ†	2	68	.78
KA—RIQ	2	68	.86
TG—RIQ	2	71	.77
TG—DIQ	2	71	.84

* All correlations are statistically significant.

† Form D was used in this part of the analysis

The correlations between I.Q.'s are very similar, none of the differences being significant. The TMcN I.Q.'s agree as well with the other tests as the others agree with one another. The coefficients are similar to those usually reported between group intelligence tests.

2. *Difficulty of the Tests.*

Means and standard deviations of various scores are shown in Table 3.

TABLE 3
Means and Sigmas of Scores on Tests and Subtests

Subtest	Form C		Form D	
	M	σ	M	σ
1	18.0	3.61	16.9	3.21
2	11.6	5.18	11.9	4.80
3	16.3	3.60	15.1	4.24
4	17.9	2.60	17.2	3.21
5	17.3	4.05	16.3	3.85
6	12.8	4.02	11.8	4.07
7	9.9	1.85	8.7	2.17
Total	103.5	18.80	97.8	19.70
MA	17.5	2.44	16.9	2.51
Standard scores	116	11.79	114	12.31

It may be seen that the two forms are distinctly comparable in difficulty, and variability. The average percentages of all

items passed are 64 and 59 for Forms C and D, respectively. The authors report average difficulty values of about 56 per cent for grades 7, 9, and 11. The higher average per cent success on Form C than on Form D for the present sample is explainable in terms of growth, with possibly some practice effect.

The subtests vary considerably in mean difficulty and variability, subtests two and six being significantly more difficult than the others.

3. *Item Difficulty.*

Form D was analyzed to determine the range of difficulty values of each item. These are shown in Table 4.

TABLE 4
Range of Percentage Success by Items

Subtest	Range of per cent success	Per cent of items between 40 and 59 per cent
1	18-98	16
2	4-92	24
3	16-98	28
4	12-100	16
5	8-98	4
6	14-95	16
7	59-96	0

With the exception of test 7, the range of success varies from a low to a high percentage in each subtest, a situation usually associated with maximum reliability (4, p. 32). On the other hand Symonds (10) and T. G. Thurstone (15) have shown that a test consisting of items of fifty per cent difficulty value measure an individual most accurately. Comparatively few of the items on this test fall within the range 40 to 59 per cent difficulty value.

The authors believe that the test is essentially a power test, i.e., that the items have been arranged within each subtest in increasing order of difficulty with ample time limits. This claim was appraised by computing the rank order correlations between obtained order and test order in the subtests of Form D. These are given in Table 5.

TABLE 5
Rho between Obtained and Test Order of Difficulty

Subtest	rho*
1	.79
2	.85
3	.87
4	.91
5	.79
6	.86
7	.69

* All values of inferred r are significant

The values of these correlations indicate that essentially the items are arranged in order of difficulty. Despite this, item analysis indicates that for the Canadian sample some items are very seriously misplaced. These results may be compared with those of Hovland and Wonderlic, who report rank order correlations between test order and obtained order of .46 to .75 in various forms of the Otis Self-Administering Test, Advanced Form (6).

There is no definite way of knowing which items a student tried, but for purposes of this analysis it was assumed that a student attempted all items down to the last one he marked. Table 6 shows the percentages of students who marked the last item in each subtest, i.e., the percentages who attempted all items.

TABLE 6
Percentages of Students Attempting All Items

Subtest	%
1	79
2	84
3	77
4	96
5	74
6	92
7	83

Evidently the test is essentially a power test, since such large numbers of subjects were able to try all items in each subtest.

4. *Suitability of the Test at the Grade 9 Level.*

The fact that about 60 per cent of all items were success-

fully passed suggests that the test might be too easy at the grade 9 level. Table 7 shows the percentages of subjects who obtained mental ages of 19 and over, and 20 and over, on each form.

Since the tests fail to discriminate between the mental ages of such a large percentage of students, and because so many earned the maximum mental age, it seems reasonable to conclude that the tests are too easy at the grade 9 level, and possibly even for bright students in grade 8. This test apparently suffers from the same weakness as did the TG test: "As Terman points out, a child capable of earning a score of 180 or better is under a handicap" (1, p. 157). A 12-year-old student may

TABLE 7
Percentages of Students with M.A.'s of 19 and 20

M.A.	Form C		Form D	
	No.	%	No.	%
19 and over	32	36	25	25
20 and over	22	25	14	16

earn a DIQ of 161, whereas the highest DIQ obtainable by an 18-year-old is 138 (11, Table 3). DIQ's are probably more satisfactory than are RIQ's, but the test appears to be too easy for students above grade 8. This should be verified by an analysis of the test results of grade 11 students.

5. *Reliability of Tests and Subtests.*

Reliability coefficients were determined by correlating scores on the equivalent forms.

The inter-form reliabilities of the subtests vary rather considerably, being .40 and .84 for subtests 7 and 2, respectively. Averaging these coefficients for the seven subtests and predicting the reliability coefficient for a test seven times as long (2, p. 283) gives an estimated reliability coefficient of .93, as compared with the obtained correlation of .94.

The correlations in the lower part of the table indicate the necessity of stating the reliabilities of all measures which teachers may use, since the various types of scores are not necessarily equally reliable (7, pp. 122-3).

TABLE 8
Reliability Coefficients

Variables	Equivalent forms
Subtest 1	.52
2	.84
3	.78
4	.58
5	.76
6	.69
7	.40
Total raw score94
Standard score90
Mental age91
DIQ89
RIQ93
Manual*96

* This is apparently based on raw scores for the age range 13-6 to 14-5, although the manual does not make this clear

Probable errors of measurement are given for certain types of scores: (a) for standard scores $P.E._M = 2.6$, compared with 2.2 reported in the manual; (b) for DIQ's: $P.E._M = 3.06$; (c) for RIQ's: $P.E._M = 3.45$.

Factor Analysis

In the manual of directions the authors state that they have chosen the content in such a way as to "have a test more highly saturated with a common factor or ability" (10, p. 1). In revising the TG test, for example, they eliminated those subtests which appeared to measure a numerical ability, so that the present revision is thought to measure "general verbal intelligence" (11, p. 1). While, of course, the number of subtests is probably too small and the reliabilities somewhat in-

TABLE 9
Intercorrelations (Form C in upper, Form D in lower part)

Subtest	1	2	3	4	5	6	7
1		.57	.50	.55	.50	.45	.36
2	.53		.67	.50	.63	.81	.56
3	.45	.64		.58	.46	.67	.48
4	.55	.50	.48		.41	.56	.49
5	.33	.43	.71	.71		.55	.53
6	.49	.85	.63	.67	.64		.55
7	.16	.46	.55	.40	.48	.49	

adequate to give a satisfactory indication of the factor loadings obtainable from these tests, a factor analysis might give some indication of the extent of the general factor. Table 9 shows the intercorrelations, those for Form C being above and those for Form D below the diagonal.

With few exceptions the intercorrelations for the two forms are of about the same magnitude. The first factor loadings and the communalities for each form were computed by the multiple-factor method (14). The obtained communalities varied somewhat from the first estimated communalities, which were taken to be the highest r in each column. This is, of

TABLE 10
Factor Loadings and Communalities for Each Form

Subtest	Form C				Form D			
	1st App.		2nd App.		1st App.		2nd App.	
	I	h^2	I	h^2	I	h^2	I	h^2
1	.67	.45	.65	.43	.59	.34	.56	.31
2	.87	.76	.87	.76	.82	.67	.80	.64
3	.77	.59	.76	.58	.80	.64	.80	.64
4	.70	.49	.69	.48	.77	.59	.76	.58
5	.71	.50	.69	.48	.77	.59	.76	.58
6	.84	.71	.83	.69	.89	.78	.89	.79
7	.68	.46	.66	.44	.59	.35	.56	.32

course, to be expected with such a small battery of tests. A second approximation was made in each case. Only one factor loading was computed since the correlations in the first residual matrix were all less than 4 times the probable error of the corresponding original correlations, making further analysis unnecessary (14, p. 26).

The results of the analysis are shown in Table 10.

It appears that little was gained by making the second approximation since practically identical factor loadings were obtained on both approximations. The factor loadings are very similar for forms C and D. In general, subtests 1 and 7 are less saturated with the common factor than is the case of the other subtests. This was verified by a cluster analysis (17), and also by the calculation of B-coefficients (5).

Since the subtests vary in their reliabilities and in their

factor loadings, the question of the possibility of shortening the test without loss arose. Subtests 2, 3, and 6 have the highest reliabilities, and the highest first factor loading. Subtests 4 and 5 have identical factor loadings but the former is less reliable than the latter. Possibly a combination of subtests 2, 3, 5, and 6 would give satisfactory results. The inter-form correlation of scores on these four subtests was found to be .92, almost as high as the reliability coefficient of total raw scores. The use of these four subtests would reduce testing time from 48 to 29 minutes, a saving of 40 per cent.

Conclusion

In general, the results of this analysis are very similar to the data reported in the manual of directions, with the criticism that the test may be too easy at the grade 9 level since it fails to discriminate between the mental ages of about 20 per cent of the present sample. The suggestion is made that the test could be considerably reduced in content with little loss in reliability.

REFERENCES

1. Carroll, H. A. *Genius in the Making*. New York: McGraw-Hill, 1940.
2. Garrett, H. E. *Statistics in Psychology and Education*. New York: Longmans, Green and Co., 1937.
3. Guilford, J. P. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill, 1942.
4. Hawkes, H. E., Lindquist, E. F., and Mann, C. R. *The Construction and Use of Achievement Examinations*. Boston: Houghton-Mifflin, 1936.
5. Holzinger, K. J. and Harman, H. H. *Factor Analysis*. Chicago: The University of Chicago Press, 1941.
6. Hovland, C. I. and Wonderlic, E. F. "Critical Analysis of the Otis Self-Administering Test of Mental Ability—High Form" *Journal of Applied Psychology*, XXIII (1939), 367-387.
7. Jackson, R. W. B. and Ferguson, G. A. *Studies on the Reliability of Tests*. Toronto: Department of Educational Research, Bulletin No. 12, 1941.
8. Lee, J. M. and Lee, D. M. *A Guide to Measurement in Secondary Schools*. New York: Appleton-Century, 1936.
9. Stuit, D. B. "Current Construction and Evaluation of Intelligence Tests." *Review of Educational Research*, XI (1941), 9-24.

10. Symonds, P. M. "Choice of Items for a Test on the Basis of Difficulty." *Journal of Educational Psychology*, XX (1929), 481-493.
11. *Terman-McNemar Test of Mental Ability*. Yonkers-on-Hudson World Book Co., 1941.
12. Thirty-Seventh and Thirty-Eighth Annual Reports of the Vancouver City Schools for the Years Ending Dec. 31st, 1939, and Dec. 31st, 1940. Vancouver, B. C.
13. Thirty-Ninth Yearbook of the National Society for the Study of Education. *Intelligence: Its Nature and Nurture*. Bloomington: Public School Publishing Co., 1940.
14. Thomson, G. H. *The Factorial Analysis of Human Ability*. New York: Houghton-Mifflin, 1939.
15. Thurstone, T. G. "The Difficulty of a Test and its Diagnostic Value" *Journal of Educational Psychology*, XXIII (1932), 335-343.
16. Traxler, A. E. "Study of the California Test of Mental Maturity-Advanced Battery" *Journal of Educational Research*, XXXII (1939), 329-335.
17. Tryon, R. C. *Cluster Analysis*. Berkeley, Calif.: Associated Students Store, University of California, 1939.

THE ROLE OF TESTS IN THE DIAGNOSIS AND CORRECTION OF SPELLING DEFICIEN- CIES OF COLLEGE STUDENTS

FRANCES ORALIND TRIGGS

American Nurses' Association

The Problem

AN EXAMINATION of the literature would seem to indicate that the college student who is a poor speller has received little encouragement to do anything about improving his spelling skills. This is in contrast to the encouragement given the college student who is a poor reader through remedial classes and clinics. The complete explanation for this situation is not clear. However, the following closely related observations may partially account for it:

1. Scientific study and diagnosis of spelling difficulties have lagged behind comparable work in reading;
2. No clear-cut and easily applicable remedial techniques in spelling have been available;
3. Teachers of college students are convinced that, if a student were ever going to learn to spell, he would have done so by the time he reached college.

There is growing evidence, however, that reading and spelling, to say nothing of other language skills, are closely related and that actually much can be done to remedy deficiencies in them even at the college level.

To that end, a remedial spelling program was set up at the University of Illinois during the academic year 1942-43. Remedial techniques were sought for this program which would require students not only to read about spelling, but also to have the experience of applying the principles studied. It was felt that by using such techniques, there was some assurance

that the students could more easily apply the skills both in and out of their class work.

For these reasons, a manual of exercises was the technique chosen. The spelling manual devised consists, first, of a discussion of spelling in general, of the ways by which spelling is learned, and of the types of skills involved in spelling; second, of a discussion of the principles of pronunciation with emphasis on those especially applicable in aiding spelling; third, of a discussion of word families; and fourth of a series of "spelling conventions" which help the students to see the system behind the spelling of many words.

Answers to two main questions were sought from the remedial spelling program: first, is it possible to improve spelling skills of college students by use of spelling exercises which require the student not only to study the principles of good spelling but also to apply them; and second, what kinds of skills and abilities must students have who may be expected to improve through this remedial technique, i.e., what background is necessary on which to build spelling skills by use of such a technique?

Procedure

Announcements were made in Rhetoric I and II notifying students that they could apply for work in the remedial spelling classes. One hundred forty-nine students applied, of whom one hundred were accepted in the first remedial sections opened. Approximately seventy students appeared at the first meetings of the classes. The work was carefully explained during this first session. Students were told that they would be required to do the assigned work and do it regularly, if they were going to attend the sessions. It was expected that every student would attend classes regularly once a week and spend at least two hours each week in preparation of manual exercises. Students were urged to come to the instructor's office for special help previous to any class period if they had difficulty doing their assignments.

Approximately twenty students did not return after this session, leaving about fifty in the four sections. Of these fifty, twenty-two were called out with the Emergency Reserve Corps.

Thus only about twenty-eight remained in the class long enough to complete the work.

Shortly before the first spelling sessions were over, a second course was arranged to accommodate those students who had not originally been accepted. This time some twenty students attended the first session, and about fourteen remained after the students understood the work which would be involved. Because of late applications, still another section was opened in which the students had to do twice as much work a week as had been planned originally. There were only about three who were able to do this. Test-retest evidence did not indicate that these students were handicapped by having to work at greater speed.¹

Certain objective test data were available on these students. Scores were available on the *American Council on Education Psychological Examination*. This is a scholastic aptitude test having two types of scores, "L," and "Q." The L-score purports to be indicative of language facility and related to the student's ability to do work requiring this type of facility, such as course work in English, foreign language, and social sciences. The Q-score purports to be indicative of the student's facility to do work requiring quantitative thinking such as is required in science and mathematics.

In addition to these scores, scores on four other tests were available: an informal spelling test of the dictation type, one of the recognition type, the *Minnesota Clerical Test*, and a phonics test. The recognition spelling test given was the spelling section from the *Cooperative English Test, Form O*, testing ability to recognize which of several spellings of a word is correct. The *Minnesota Clerical Test* has two sections, one on names and one on numbers. The numbers section consists of columns of numbers in pairs. If the pair is exactly the same, the subject is to check it. The names section of the test is similar. This test is closely timed and thus requires both speed and accuracy. The phonics test has two parts. Part I tests the student's ability to divide words into syllables; Part II tests his ability to sound words according to a somewhat

¹ No check on extent of comparability of these two groups was made.

simplified arrangement of the usual dictionary key to pronunciation.

Scores from these tests and from certain clinical data indicated to some extent the types of difficulties which individual students had. There were those students who had good basic language abilities and skills as shown by a high score on the "L" of the *American Council on Education Psychological Examination*, names of the *Minnesota Clerical*, and the usage and vocabulary subtests on the English test but low scores on the spelling subtest and the phonics tests. The problem in such a case seemed to be to make the student aware of the need for accurate spelling and to show him how to apply his skills by any of several techniques.

The tests also revealed those students who had a potential facility in language, but who had never developed language skills. There were also those students who probably do not have the general ability and potentialities to develop the language skills necessary to succeed in college.

Students attended class for eight weeks for one hour a week. Each student was given in dittoed form spelling exercises from the manual described earlier (Frances Oralind Triggs and Edwin Robbins, *Improve Your Spelling*, New York: Farrar and Rinehart, 1944). Individual conferences with students allowed the instructor to individualize somewhat the work in the manual to fit student needs as shown by the informal diagnoses made from the type of work done both in and out of class.

Those students who cared to take retests were given different forms of the same tests which they had taken at the beginning of the work. These tests were then interpreted for them in individual conferences. There are two types of interpretation which can be made from such retests: interpretations which apply to individuals only and interpretations applying to the group as a whole. Individual interpretations are mainly of value in guiding the further growth of the student, and are made on the basis of both an intimate knowledge of that student and experience with the group as a whole. Group interpretations show general trends which result from remedial work. Both serve as a basis for evaluation and modification of procedures.

Results of Remedial Work

The dictation spelling test was built to illustrate the principles discussed in the manual. The students at no time studied the specific words in the test. Test-retest evidence indicates a range from a gain over the remedial period of ten words to a loss of one word, with a mean gain of 3.6 words. It is probable that results from this type of test most nearly reflect the ability of the student to do the spelling task usually required of him.

Test-retest evidence on the clerical test was interesting in that there was a greater gain on the names section than on the numbers section. The range of gain on the names part was from 36 to 0, with a mean gain of 17 words. The range of gain on the numbers section was from 33 to minus 14, with a mean gain of 13 items. This group of students originally had markedly higher scores on the numbers section of the test than they had on the names section. On the retests, this difference was not so evident. Gains on this test probably indicate an improvement in ability to look within the word and recognize word parts rather than in ability to recognize the word only by its configuration. It is this type of skill which is used in proofreading and in reading where it is necessary to distinguish between words of like configuration such as "physiology" and "psychology," "insulation" and "installation," and in many cases such simple words as "then" and "than," "also" and "solo," and others. This type of skill probably should not be over-emphasized because it might adversely affect reading skills. However, a balance between work of this kind and work on skills required in normal silent reading will probably result in improvement in both reading and spelling.

Gains were also evident on the phonics test. On the syllabification section, Part I, the mean gain was ten words, with a range of from 22 to minus two. On Part II of this test, the ability to sound words, the range of gain was from 27 to minus four, with a mean gain of 11 words. A gain on this test, when accompanied with gain on a spelling test, suggests that students not only have learned the tools of word recognition but also are beginning to apply them. When these same skills

were measured by oral reading, it became even more evident that students not only had learned them but actually were putting them into practice.

The Reaction of the Students to Remedial Work

It was interesting to note the reasons students gave for registering in this course. In terms of the stated motives of the students, they might be classified as follows: First, there were the students who were merely curious to know what the work would be like, but who did not care to put time on remedial work. Second, there were the sincere students who wanted to improve their spelling skills, but who actually did not have the time available to work through the manual. Many of these students were carrying heavy schedules besides actual work to help finance their education. This type of student is the one who is most severely handicapped by poor verbal skills. Our university curriculum requires a great deal of verbal work, yet it takes these students who have poor verbal skills longer to do the work; therefore, they do not have the time to put on the remedial work, and the longer they spend on their class work, the less chance there is that they will be able to put in the extra time on improving their skills. This is an illustration, surely, of the old saying "them that has, gets." Third, there was a group of very sincere students who had time to do the work, and who did excellent, consistent class work. Some of these students were handicapped by poor scholastic aptitude and did not gain as much in the end as their efforts warranted; but most of this group made excellent improvement as measured by both daily written work required in their courses and by standardized tests.

At the four weeks' point in the remedial work, to remind the students of the importance of consciously trying to transfer skills learned in their remedial work to class work, the instructor asked the students to write during class time an informal five-minute essay, expressing their reaction to the remedial work, and indicating whether they had been able to notice any improvement in their spelling up to that time. A number of the reactions, written both at this time and later, are given below.

The real purpose of this letter is to thank you for your help. The value of your spelling course showed itself clearly in my last theme for Verbal Expression. Though it was one of the longest compositions, it contained fewer errors than any preceding paper. I misspelled only two or three words. It seemed almost unbelievable to turn page after page without an error.

It will, as you said, be some time before I am able to realize the full benefit of your instructions. But already I can print legibly and at a reasonable speed. My spelling is improving, and one may see something in my way of doing things which resembles organization. My enunciation (thanks to your advice to visit the speech clinic) has shown some improvement. It will continue to develop since now I have the rudiments and need only practice.

All these things you've done for me against my own objections. It would have been easy for you to let me go when I was determined to give up. It was some time before I could appreciate this work of yours. Now I can see what it has done and will do, so I want to apologize for my lack of character, and thank you for all you've done for me.

I have been a student of the experimental remedial spelling course for the past four weeks. In that time there has been a slow transition of confidence within me in all phases of handling and working with the English language. This change may not be outwardly apparent at this present moment, but I'm sure time will bear out that there is a *definite* improvement in this respect.

My one regret, in regard to this course, is that it is of only eight weeks in length.

From remedial spelling I have received an improvement in spelling. I have never studied related words before or paid much attention to the way the words were pronounced. These simple things have aided my spelling. Before I took this course I never thought of the different ways of spelling words—hand, ear, etc.

When I started to the University of Illinois I was very weak in spelling. In fact I don't think I could have been much worse. It seemed that I just couldn't learn to spell. I couldn't find out what was the matter. I was offered a chance to take this extra spelling course to improve my spelling. I was very much pleased with the chance, so I enrolled in the course. I have just finished four weeks of the eight-weeks course and I am beginning to see more closely some of the basic fundamentals of spelling which I had completely missed before. I can't say after four sessions that I am an

outstanding speller, but I do believe that I will be a better speller after I have finished the course.

I know that the four hours which I have spent in spelling class have helped me a great deal. I have been doing better work in my regular English class and the letters which I have sent home have improved.

I still am a very poor speller; however, I am able to find some of my faults. I believe before the spelling classes are over my spelling will improve a great deal more than during the first four weeks.

I believe these spelling classes should be given next year so other students may also have a chance to improve their spelling.

I believe that the help I am getting from our spelling class will not only help me to overcome spelling troubles, but it will be a great help in obtaining exactness with all my other work as well. In fact I have already been helped by the principal parts which we have taken up, mainly forming a picture of the word I am hunting for. Yesterday, for example, I had to write a theme about myself while I was in the process of being sworn in as a Naval Cadet, and I was bothered with the spelling of a couple of words I chose to use. My sight spelling came to my rescue, and I was able to do a decent piece of work on my theme. This is only one instance that I remember because it was so recent and much depended on it.

I'll admit that after the first few classes of remedial spelling, and after seeing the long and seemingly difficult assignments I was disgusted with myself for enrolling. I had always told myself that I was almost infallible in spelling, but my mother was very disgusted about the lack of phonics in our grade-school system and insisted that I was a poor speller. Spelling came easy for me and I imagined that remedial spelling in college would be one continual spelling match, and they are fun. However, I found that the accuracy the work requires is helping me in many ways. I've discovered that there are many facts about spelling I had never thought about. I believe that this remedial work should be included in college Rhetoric and English sections, because many freshmen, newly graduated from high school, lack the fundamentals, training, and background to spell correctly, and the thoroughness of the work and assignments will aid in every course.

Reactions of the Faculty to Remedial Work

The faculty of the English Department was, at all times, aware of what was being done in the remedial spelling work.

They referred students to it, and twice the work of the remedial spelling classes was described at English staff meetings. The cooperation of the faculty with the instructor in remedial spelling was excellent. There are many indications that the instructors welcomed the special help given these students. Many of them felt that they had very little time to give individualized help in spelling, but that if such could be done, the results would be worth while. Certain comments of the faculty on individual cases are given below.

Thank you for your note about Mr. X. He has spoken to me of the excellent help you have given him with his handwriting and with his spelling. I am greatly pleased with his progress, and with his attitude toward you personally. I shall be referring students to you in the future, urging them to take advantage of the opportunity of following your suggestions.

Your course in remedial spelling has been of considerable help to my student, Jack Doe. Originally, he was by no means a hopeless speller; but his spelling was bad enough to handicap him in his work. The carelessness and the word-ignorance which caused many of his errors have been checked, I think, by the work he has done with you. On his themes, at least, he has shown an increasing awareness of the necessity of correct spelling. Part of his improvement has come, no doubt, from his general development in language skills as a whole, through his work in Rhetoric, and from his own intellectual and social growth; but your work with his spelling has unquestionably given him valuable help with that particular aspect of his training.

Mr. Doe has not, of course, been suddenly transformed into a perfect speller. That is too much to expect. But he has developed an interest in words themselves and has come to realize the importance of *thinking* while spelling. It is this new attitude, I think, which will have the most bearing on his continued improvement in spelling.

If other students have gained from their work in remedial spelling as much as Mr. Doe has gained, I think the course certainly should be continued for the benefit of future students.

You asked what results your spelling class had on my student, Mr. X.

To begin with, his spelling was very bad, though largely, I think, through carelessness. Almost at once his home themes showed great improvement as he became more con-

scious of his needs, and by the end of the semester, he rarely missed more than one or two—usually easy—words in each of them.

If that improvement lasts, and if your other students did as well, I should certainly want to see the class continued.

I am pleased with the progress made in spelling by Miss Blank and Mr Long. The spelling grades alone do not evaluate the counsel and assistance you have given these students. Your remedial spelling is a worth-while project and should be continued.

Conclusion

The major generalizations to be drawn from this study is that poor spellers can improve their spelling skills by a remedial technique such as has been described. This generalization can be made more specific by some further comments.

There are rather complete records for ninety of the students of this group. A study of these records indicates the importance of careful attention to the reasons for the spelling difficulties. For instance, twenty-six students had poor spelling skills mainly because of carelessness, lack of the habit of proof-reading what had been written, and, in general, an attitude that spelling is unimportant. Sixty-four students, however, lacked at least some of the following skills: They could not divide words into syllables, nor could they accent words correctly. They had very little knowledge of the construction of words—that is, they did not know what suffixes and prefixes were. They did not realize what base words or root words were—and when reading orally they miscalled words of like configuration. Thus it became evident that they had no methods for attacking new words. They also had little knowledge of spelling “conventions.” Many of this group were not only poor spellers, but poor readers; and many of them had poor English skills as measured by the objective test given them at the beginning of the year and by subsequent class work.

On examining these records, it is possible to make a prognosis of the extent of success of these students as the result of remedial spelling work if general ability is taken into account. In this regard, it might be said in general that if

there is some indication of measured general ability and if remedial work follows a careful diagnosis of difficulties, the prognosis of success in remedial work will be good, assuming the student applies himself assiduously. But for the student who does not have measured general ability, successful results cannot be universally predicted. However, it is always possible that a student's poor scores on the general ability test may be due to lack of development of language skills. If there is time available, an individual ability test can be used to determine to what extent the student is penalized by the form of the test given. If on the basis of an individual test potential ability is evident and if plenty of time is available for remedial work, satisfactory results may be forthcoming.

Probably the major error made in this remedial program was that it was placed, for most students, on top of an already over-full schedule. Requirements of the remedial program were heavy. These students are already the ones who have to spend the most time in the preparation of their courses because of lack of verbal facility, which is a greatly needed tool throughout the university curriculum.

Recommendations

On the basis of experience with this remedial program, it is recommended, first, that the students who are poor spellers be segregated and their records examined at the very beginning of the school year; second, that the reason for this disability be determined in each individual case; third, that a stated requirement be made of these students if they are to pass English; and fourth, that a special place in the curriculum be given for remedial training as may be required. If the student's disability is great enough, his whole program should be lightened to allow time enough to do the remedial work, and do it well. It has been found time and again that, where such an approach is taken, the student's improvement is apparent not only in spelling but in other language skills as well, and that this improvement is carried over into his course work.

One further observation should be made. The motivation of the student is a major factor in the degree of success he will

have in any type of remedial work but should probably receive special consideration in the remedial spelling program. The extent to which it is important for an individual to follow spelling conventions will probably be a determining factor in his motivation for remedial work in spelling. Though the clinician or instructor working with him may realize that probably no strict line of demarcation exists between reading, spelling, and other language skills, it may be difficult to convince the student of this fact. If he is aware only of his spelling disability and has "gotten by" this long, it may be somewhat difficult to convince him that he cannot always "get by" with no handicap to himself. It is therefore recommended that the well-motivated students, as well as the students for whom prognosis of success in remedial spelling is good, be the ones to receive attention first, at least while remedial techniques are being evaluated.

There is always the question of how much responsibility the university can take in developing sub-college English, spelling, and reading skills. This, of course, is a matter of policy to be set by the school in question. However, it is suggested that, if it is possible to demonstrate that spelling can be taught at the college level, the public schools may be helped to realize that it can also be taught at the lower educational levels. They may then take over the responsibility at that level and relieve the college of the necessity of worrying about it.

DISCRIMINATIVE VALUE AND PATTERNS OF THE WECHSLER-BELLEVUE SCALES IN THE EXAMINATION OF DELINQUENT NEGRO BOYS

JOSEPH CHARLES FRANKLIN¹

Civilian Public Service # 115, Laboratory of Physiological Hygiene,
University of Minnesota

THE psychologist working with delinquents in an institutional setting is obliged usually to maximize the validity and utility of his findings in individual case and group studies with the least expenditure of time, energy, and resources. Consequently, he is most likely to turn to the supply of available tests and, applying criteria growing out of his purposes and determining test "goodness" in relation to prospective testees, to select those test materials which are most easily administered, scored, and interpreted.

In intellectual measurement the use of tests on subjects differing from the standardization populations from which the norms derive, in one or more significant variables, involves concern with the attainment of valid and meaningful measurement.

The Cheltenham School for Boys, Cheltenham, Maryland is a State Institution for delinquent Negro boys. The background of the boys committed is commonly one of social and/or personal maladjustment. Their previous life conditions are marked by broken homes, inadequate familial organization and integration, poor supervision, and neglect. The incidence, variously, of sub-standard shelter, poverty, lack of medical care, and even malnutrition, is preponderant. These children are seriously retarded educationally, approximately at the

¹ The writer is indebted to Donald L. Grummon for cooperation in the administration of the *Scales* and to Charles W. Piersol for assistance in the completion of the study.

third-grade level, at the chronological age of fourteen and a half; and truancy, suspension and expulsion from school, and consistent failure largely characterize their formal education.

At the practical level of mental testing of such subjects, awareness of and consideration for the implications of the state of psychological knowledge in such areas of theoretical research as the following are of methodological and evaluative importance: nature and nurture, race and nationality differences, rural and urban effects, equality of normal opportunity for socially, educationally, and intellectually stimulating experience or the lack of it, the fixity or flexibility of mental capacities, and the organization of mental abilities.

It is beyond the scope of this study to discuss the relationships between the conflicting conclusions of research in these fundamental problems and the construction and use and interpretation of obtained results in the mental measurement of Negroes. Highly useful references are provided in bibliographies compiled by Bean (1) and the editors of the *Journal of Negro Education* (14).

Nevertheless, keeping the relevance of the basic issues in mind serves two worth-while purposes. First, survey of available tests reveals the inadequacies of existing materials with resultant difficulty in selecting a "good" test (particularly with regard to standardization and norms) for Negroes, much less delinquent Negro children. Secondly, the need in intellectual measurement is observed to be shifting from simple over-all characterization of mental status to intra- and inter-individual comparisons of partialled-out components or aspects of intellectual functions. It becomes obvious that in these terms tests easily administered, quickly scored, readily interpretable, and suitable to our subjects are not available.

Preliminary use and appraisal of various group and individual mental tests were made. It was found that the *Wechsler-Bellevue A & A Scales* provided maximally useful information regarding mental status and facilitated needed qualification of test results with respect to the fundamental problems already mentioned.

Wechsler did not include Negroes in his standardization

and specifically urges caution in the use of his test with non-whites. Nevertheless, the apparent and distinct advantages of the *Wechsler-Bellevue Scales* in classification, insofar as classification depends upon mental status, warranted further use and study of the test. The bias of the standardization as related to this minority group is a serious incongruity but substantially no greater than that involved in the use of other tests the results of which did not favorably compare in usefulness and meaningfulness with the *Wechsler-Bellevue*. The proper extension of the use of the *Wechsler-Bellevue* to Negroes depends upon such data as those which this report in part provides.

Purposes of the Investigation

In order to assess objectively the suitability of the *Wechsler-Bellevue Scales* for the intellectual testing of institutionalized Negro boys, this study was undertaken to answer the following questions: How does the test sift and sort the population as to mental level? Do the sub-tests positively discriminate among the subjects as they are classified within the various mental level categories? What are the patterns and trends of performance of the total and sub-groups on the sub-tests? Is the suggested use of a short form warrantable with this population?

Procedure

Two hundred and seventy-six boys were given the *Wechsler-Bellevue* (both Verbal and Performance Scales) during 1943-44. The average institutional population during this period was about two hundred and seventy. For the most part boys were routinely tested shortly after admittance but some were especially referred for testing for purposes of classification from among those admitted prior to the initiation of the program of intellectual testing.

The *Wechsler-Bellevue Scales* consist of eleven sub-tests, one of which is the Vocabulary alternate in the Verbal Scale, which was not used. The five Verbal sub-tests depend heavily upon language for administration and for subject responses. These primarily involve abstractual, conceptual, and general-

izing mental functions. According to Wechsler as reported by Rosenzweig, Bundas, Lumbry and Davidson (10) they are described as follows:

1. *Information*: consists of questions formulated to tap the subject's range of information on material that the average person with average opportunity should be able to obtain for himself.

2. *Comprehension*: measures the use of "common sense" and judgment in situations described to the subject. Success on this test seemingly depends upon the possession of a certain amount of practical information and a general ability to use past experience.

3. *Arithmetical Reasoning*: measures mental alertness as well as ability to handle practical calculations.

4. *Memory Span for Digits*: measures immediate memory for digits forward and backward.

5. *Similarities*: measures ability to discriminate between essential and superficial likenesses; to generalize and think in abstract terms.

The five Performance sub-tests require the subject to manipulate concrete materials and to perform certain tasks such as arranging pictures and assembling object forms. The same authors describe them as follows:

6. *Picture Arrangement*: detects ability to comprehend or "size up" a total situation.

7. *Picture Completion*: measures ability to differentiate essential from unessential details.

8. *Block Design*: a test of general intellectual functioning, involving both synthetic and analytic ability, but weighted considerably with ability to solve problems in spatial relations.

9. *Digit Symbol*: measures speed and accuracy of learning new associations.

10. *Object Assembly*: measures insight into spatial relationships of familiar objects.

Each sub-test contains items which are related to a component mental function and the items are arranged in order of increasing difficulty. Scores on sub-tests are converted into "weighted" scores which make possible direct comparison of the various sub-test performances. Separate Verbal and Performance I.Q.'s are obtained by summing the appropriate sub-tests, and these in turn are combined in an over-all measurement, the Full I.Q.

Results

The chronological age range of the 276 subjects was from 9.63 years to 20.13 years with a Mean of 14.6 and a S.D. of 1.56 years. Results for the total group are given in Table 1.

TABLE 1
Average Performance for Entire Group (276 Cases)

	Mean	Median	S.E. Mean	S.D.
Full I.Q.	76.5	76.6	.926	15.39
Verbal I.Q.	76.2	75.8	.869	14.45
Performance I.Q.	80.4	82.9	1.094	18.19

On the basis of individual test results the subjects were grouped according to Wechsler: Normal (91-110); Dull Normal (80-90); Borderline (66-79); Mentally Defective (below 66). Test results for these groups are presented in Table 2. Comparisons of measures of central tendency and dispersion may be made since the age distributions within the sub-groups are practically identical. These results are summarized in Table 3.

TABLE 2
Performance Data of Sub-Groups According to Mental Level

Group	N	Mean I.Q.	Median I.Q.	S.D. Mean
Normal	52	98.0	97.3	5.29
Verbal	52	95.3	95.4	7.12
Performance	52	100.8	100.8	6.86
Dull Normal	64	85.1	84.7	3.37
Verbal	64	82.6	82.4	7.26
Performance	64	90.1	89.5	6.05
Borderline	90	72.9	73.4	3.94
Verbal	90	73.1	72.9	7.17
Performance	90	78.8	79.5	7.62
Mentally Defective	70	55.8	56.0	7.53
Verbal	70	60.8	60.0	8.20
Performance	70	60.7	60.3	10.62

Discriminative Value of the Sub-Tests

Wechsler, Israel, and Balinsky (13) and Lewinski (5) have reported positive discriminative values of the sub-tests of the *Wechsler-Bellevue Scales* in differentiating between the various

TABLE 3
Age Data for Mental Level Sub-Groups

Group	N	Range	Mean age	S.D Mean age	S.E Mean age
Normal	52	11.13-18.63	14.8	1.63	.218
Dull Normal	64	9.63-18.63	14.5	1.64	.208
Borderline	90	11.13-18.63	14.4	1.34	.141
Mentally Defective . . .	70	10.11-20.13	14.6	1.64	.196

intellective levels. Their studies, however, were done with quite different samples from that with which we are here concerned.

In order to ascertain the discriminative values of the sub-tests in differentiating between subjects categorized on the basis of total test results, the differences in mean weighted scores, the standard errors of these differences, and the critical ratios were calculated. Table 4 shows that all of the sub-tests discriminate between the various levels with the exception of three: (1). The Digit Span did not satisfactorily distinguish the Normal from the Dull Normal subjects, (2) the Digit Symbol did not significantly discriminate the Dull Normal from the Borderline, and (3) the Picture Arrangement did not significantly separate the Normal from the Dull Normal. While the results generally agree with those of Wechsler, Israel, and Balinsky and with those of Lewinski, they differ at several points. The former found the Digit Span test of questionable value in discriminating between Borderline and Defective subjects whereas in this situation the same test does discriminate significantly between these two groups. The latter obtained significant discrimination on the Digit Span between all groups. In this study, however, the Digit Span failed to differentiate significantly between the Normal and Dull Normal groups.

Patterns of Sub-Test Performance

Inspection of the sub-test performances (see Table 5) shows that for the entire 276 subjects the five best-performed were in the Performance Scale with the exception of Block Design, the Similarities test of the Verbal Scale placing fourth in the list of the first five. Accordingly, Block Design plus all of the

Verbal sub-tests with the exception of Similarities ranked in the lower half of the ten sub-tests. In rank order the three highest, i.e., best-performed, sub-tests were Object Assembly,

TABLE 4
Discriminative Values of Sub-Test Performances Between Mental Levels

Sub-test groups		Difference	S.E. Difference	C.R.
1	Information			
	Normal—Dull Normal	2.50	.15	16.6
	Dull Normal—Borderline . . .	1.17	.26	4.5
	Borderline—Mentally Defective	1.14	.17	6.7
2	Comprehension			
	Normal—Dull Normal	2.75	.46	6.0
	Dull Normal—Borderline	1.05	.28	3.8
	Borderline—Mentally Defective	2.21	.23	9.6
3	Arithmetic Reasoning			
	Normal—Dull Normal	2.55	.45	5.7
	Dull Normal—Borderline	1.33	.42	3.2
	Borderline—Mentally Defective .	1.91	.37	5.2
4	Digit Span			
	Normal—Dull Normal75	.44	1.7
	Dull Normal—Borderline	1.16	.37	3.1
	Borderline—Mental Defective . . .	1.75	.35	5.0
5	Similarities			
	Normal—Dull Normal	1.73	.42	4.1
	Dull Normal—Borderline	1.12	.30	3.7
	Borderline—Mental Defective	2.67	.31	8.6
6	Picture Completion			
	Normal—Dull Normal	1.81	.43	4.2
	Dull Normal—Borderline	1.29	.39	3.3
	Borderline—Mentally Defective . .	2.82	.39	7.2
7	Picture Arrangement			
	Normal—Dull Normal	1.15	.42	2.7
	Dull Normal—Borderline	1.44	.40	3.6
	Borderline—Mentally Defective . . .	3.33	.35	9.5
8	Object Assembly			
	Normal—Dull Normal	1.36	.45	3.0
	Dull Normal—Borderline	1.44	.42	3.4
	Borderline—Mentally Defective	3.25	.44	7.4
9	Block Design			
	Normal—Dull Normal	2.35	.42	5.6
	Dull Normal—Borderline	1.80	.35	5.1
	Borderline—Mentally Defective . .	2.20	.30	7.3
10	Digit Symbol			
	Normal—Dull Normal	1.62	.34	4.8
	Dull Normal—Borderline28	.26	1.1
	Borderline—Mentally Defective . . .	1.77	.27	6.6

Picture Arrangement, and Picture Completion; the three lowest, i.e., most poorly-performed, were Arithmetic, Information, and Block Design. Quite clearly, performance materials are

TABLE 5

Data and Rankings in Performance on Sub-Tests of Mental-Level Groups

Sub-test group	Ranking of sub-test	N	Mean weighted score	S.D.	S E. mean
1. Information					
Normal	9	52	7.42	2.87	.40
Dull Normal	10	64	4.92	1.84	.23
Borderline	10	90	3.75	1.12	.12
Mentally Defective . .	9	70	2.61	1.00	.12
Total	9.5	276	4.41	2.38	.14
2. Comprehension					
Normal	4	52	9.54	2.81	.39
Dull Normal	6	64	6.79	2.01	.25
Borderline	6	90	5.74	1.21	.13
Mentally Defective . .	7	70	3.53	1.57	.19
Total	6	276	6.15	2.88	.17
3. Arithmetic Reasoning					
Normal	8	52	7.73	2.15	.29
Dull Normal	9	64	5.18	2.59	.32
Borderline	9	90	3.85	2.58	.27
Mentally Defective . .	10	70	1.94	2.22	.26
Total	9.5	276	4.41	3.12	.19
4. Digit Span					
Normal	10	52	7.48	2.35	.33
Dull Normal	7	64	6.73	2.28	.29
Borderline	7	90	5.57	2.17	.23
Mentally Defective . .	4	70	3.82	2.28	.27
Total	7	276	5.76	2.61	.16
5. Similarities					
Normal	6	52	9.32	2.55	.35
Dull Normal	4	64	7.59	1.87	.23
Borderline	4	90	6.47	1.87	.19
Mentally Defective . .	5	70	3.80	1.98	.24
Total	4	276	6.59	2.79	.17
6. Picture Completion					
Normal	3	52	9.71	2.11	.29
Dull Normal	3	64	7.90	2.53	.32
Borderline	3	90	6.61	2.11	.22
Mentally Defective . .	6	70	3.79	2.69	.32
Total	3	276	6.78	3.35	.20
7. Picture Arrangement					
Normal	2	52	10.40	2.00	.28
Dull Normal	2	64	9.25	2.47	.31
Borderline	2	90	7.81	2.33	.25
Mentally Defective . .	3	70	4.48	2.12	.25
Total	2	276	7.79	3.16	.19
8. Object Assembly					
Normal	1	52	10.73	2.30	.32
Dull Normal	1	64	9.37	2.55	.32
Borderline	1	90	7.93	2.55	.27
Mentally Defective . .	1	70	4.68	2.93	.35
Total	1	276	7.97	3.52	.21

TABLE 5 (*Continued*)

	Sub-test group	Ranking of sub-test	N	Mean weighted score	S.D.	S.E. mean
9	Block Design					
	Normal	5	52	9.38	2.28	.32
	Dull Normal	5	64	7.03	2.19	.27
	Borderline	8	90	5.23	2.09	.22
	Mentally Defective ..	8	70	3.03	1.76	.21
	Total	8	276	5.90	3.03	.12
10.	Digit Symbol					
	Normal	7	52	8.25	1.97	.27
	Dull Normal	8	64	6.63	1.66	.21
	Borderline	5	90	6.35	1.44	.15
	Mentally Defective ..	2	70	4.58	1.87	.22
	Total	5	276	6.33	2.09	.13

more efficiently handled and at a higher level than verbal materials. The results pertaining to the performance of the entire group on the sub-tests together with rank order of each of the ten sub-tests are set forth in Table 5.

For the purposes of ascertaining the patterns of performance for each of the various mental-level groups the mean weighted scores and their standard deviations on each of the sub-tests were computed. The data are tabulated in Table 5 and presented graphically in Figure 1. The striking similarity of the curves for all groups—regardless of intellectual status—indicates systematic and consistent variations for the population in organization of mental abilities and hence in their development.

For the population and for all mental-level groups the background of general information and the mental alertness linked with the ability to perform mental mathematical computations constitute a special deficiency (Information and Arithmetic). The subjects were uniformly better able to comprehend or "size up" total situations than to distinguish between essential and unessential details and parts of common objects and forms (Picture Arrangement and Picture Completion). Characteristically low performance on Block Design indicates poor synthetic and analytic abilities in dealing with more complicated problems of spatial relationships as contrasted with ability to solve problems of simple spatial relationships in assembling

familiar objects for Object Assembly, which was the best-performed sub-test in all groups.

Some differences, however, are noted in the profiles in Figure 1. Information is consonantly low but exceeds Arithmetic at the Defective level, falls at about the same place at the Borderline level but falls below Arithmetic at the Dull and Normal levels. Digit Span exceeds Arithmetic at the lower three levels but lies below Arithmetic at the Normal level.

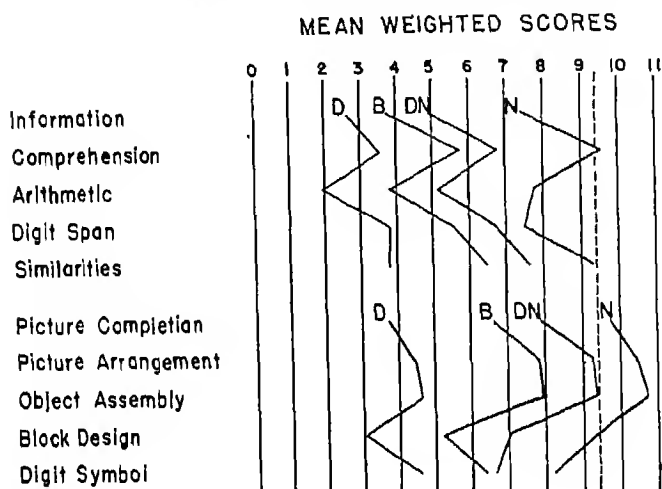


FIGURE 1

Legend - - - Average sub-test performance required to obtain Full IQ of 100 (14-6 yrs.)

D Mentally Defective

B Borderline

DN Dull Normal

N Normal

The Similarities sub-test exceeds all other Verbal sub-tests with the exception of Comprehension at the Normal level, which is higher. Digit Symbol is about the same as Object Assembly at the Defective but falls far below the latter at all other levels.

Scatter—variability of performance achievement among the sub-tests—in the *Wechsler-Bellevue* is associated with states of maladjustment, neuroticism, and psychoses. Diagnostic clinical signs are related to patterns of sub-test success and failure (3, 6, 9, 11, 12). Work in this area is in the experimental

stage and the findings reported while not conclusive as to relationships between psychometric test patterns and mental illness are suggestive. It is a matter of conjecture as to what extent psychopathology or psychological maladjustment influenced the range and level of sub-test performances of the subjects. It may be presumed, since few of the subjects examined could have been regarded as psychotic, that presence of clinical factors does not seriously mitigate against interpretation of the data according to organization and level of the mental abilities. It is noteworthy, nevertheless, that examination of the test profiles of groups above the Defective level discloses that in sub-test performance five relate positively, two negatively, and two indecisively with Wechsler's (12) diagnostic pattern for adolescent psychopathic personality trends.

The consistent and paralleling variation in sub-test performance of all subjects regardless of mental level raises important questions relevant to (1) the study of race differences in intellectual abilities and (2) the relationships of systematic lower-level performance in tests of intelligence by minority groups to the extent to which success depends upon such factors as education, training, and experience (4, 7, 14). It may be that the group patterns of sub-test performance reported here reflect relative handicaps in mental development rather than manifest strengths and weaknesses of intellectual functions. Fewer or other depressants to maximal mental development may exist in the white population on which the *Wechsler-Bellevue Test* was standardized. Investigation is needed to discriminate the sub-tests in terms of the degree to which educational and social experiences and achievements are prerequisite to differential success in sub-test performance.

Use of the Short Form of the Wechsler-Bellevue

Rabin (8) has offered an abbreviated form of the *Wechsler-Bellevue Scales*. Using the Comprehension, Arithmetic, and Similarities sub-tests and computing the total weighted score by dividing the sum of the weighted scores of these three sub-tests by three and then multiplying by ten, Rabin reported correlations of .95 with the results from administration

of the ten sub-tests. It was his opinion that the regional and educational homogeneity of his subjects rendered his choice of sub-tests a good one for a short form of the *Wechsler-Bellevue Scales*. The author stated that because the *Short Form* is primarily a verbal test it might not prove satisfactory for use with persons with a non-English language background. Rabin advised further use of the suggested *Short Form* with other groups of subjects for experimental purposes.

In order to investigate the suitability of the use of the *Short Form* with our subjects, the data were analyzed according to Rabin's method. All subjects were native-born with a common English linguistic background.

According to this method I.Q.'s differed significantly from those deriving from administration of the ten sub-tests for all

TABLE 6
Comparison of Results: Short Form and Full Wechsler-Bellevue

Group	N	Mean I.Q. full test	Mean I.Q. Rabin	Mean Diff.	S.E. Diff.	C.R.
Normal	52	98.0	97.2	- .8	1.41	.006
Dull Normal	64	85.1	80.0	- 5.1	1.18	4.3
Borderline	90	72.9	68.5	- 4.4	1.02	4.3
Mentally Defective .	70	55.8	51.4	- 4.4	1.08	4.1
Total	276	76.5	72.3	- 4.2	.58	7.4

mental-level groups with the exception of the Normal. For the total of two hundred and seventy-six cases the Mean I.Q. yielded by the short form was 72.3, which was significantly lower by 4.2 I.Q. points than the Mean I.Q. (76.5) derived from administration of the full test. In every mental-level group the *Short Form* resulted in a lower I.Q. than the ten sub-tests. In Table 6 data pertaining to the analysis are given.

It is concluded, therefore, that the use of the Rabin *Short Form* of the *Wechsler-Bellevue Scales* is not a steady or satisfactory substitute for the ten sub-tests of the *Wechsler-Bellevue* with the subjects examined. Caution dictates that the *Short Form* should not be used with subjects resembling those examined in this study. It appears obvious that the *Short Form* should not be used in the mental examination of subjects whose verbal abilities are inferior to their performance abilities.

Summary

The *Wechsler-Bellevue Scales* for individual mental testing were administered to 276 institutionalized delinquent Negro boys. The chronological age range was from 9.63 years to 20.13 years, with a Mean of 14.6 and a S.D. of 1.56 years.

The study was undertaken in order to report the results of the use of the *Wechsler-Bellevue* on this population, to investigate the discriminative values of the ten sub-tests of the *Scales* among the various mental levels, to summarize the trends and patterns of sub-test performances of the population and of the subjects grouped according to level of intellectual ability, and to examine the suitability of a suggested *Short Form* of the *Wechsler-Bellevue Scales* for the mental measurement of institutionalized delinquent Negro boys

1. Results of the administration of the *Wechsler-Bellevue* placed 19 per cent at the Normal level, 25 per cent at the Dull Normal, 33 per cent at the Borderline, and 23 per cent at the Defective level.²

2. With the exception of the Defective group, the Performance I.Q.'s exceeded the Verbal I.Q.'s by 5.5 points for the Normal group, 7.5 points for the Dull Normal, and 5.7 for the Borderline group. Over-all, the Mean Performance I.Q. exceeded the Mean Verbal I.Q. by 4.2 points.

3. The sub-tests of the *Wechsler-Bellevue Scales* discriminate significantly between the several intellectual levels (as derived from the full test) with the following exceptions: Digit Span did not prove satisfactory in distinguishing between the Normal and Dull Normal subjects, Digit Symbol between Dull Normal and Borderline subjects, and Picture Arrangement between the Normal and Dull Normal.

4. There is marked similarity in the patterns of performance from mental level to mental level. The group as a whole shows striking disparity of achievement on the sub-tests. These differences in performance have relevance to the study of racial differences. Those sub-tests characteristically per-

² A considerable increase of percentages in the higher mental levels would result if greater weight were attached to Performance achievement at the expense of Verbal in determination of the Full I.Q.'s.

formed at lower levels should be studied further in order to evaluate the role played by previous life conditions in their successful performance.

5. Consideration in interpretation of the reported results should be given to the fact that non-whites were not included in the standardization of the *Wechsler-Bellevue Scales*. Some uncalculated error of measurement may have resulted from the presence in the subjects of states of negative adjustment, of which there are indications according to the positive clinical signs developed by Wechsler and others.

6. The *Short Form* of the *Wechsler-Bellevue* by which the Full I.Q. is derived from performance on three of the ten subtests (Comprehension, Arithmetic, and Similarities) was not suited to mental measurement of the individuals examined. Evidence indicates that the *Short Form* should not be used with individuals whose Verbal abilities are inferior to their Performance abilities

REFERENCES

1. Bean, K. L. "Negro Responses to Verbal and Non-Verbal Test Materials" *Journal of Psychology*, XIII (1942), 343-353.
2. Bijou, S. W. and McCandless, B. R. "An Approach to a More Comprehensive Analysis of Mentally Retarded Pre-Delinquent Boys." *Journal of Genetic Psychology*, LXV (1944), 147-160.
3. Gilliland, A. R., Wittman, Phyllis and Goldman, M. "Patterns and Scatter of Mental Abilities in Various Psychoses." *Journal of General Psychology*, XXIX (1943), 251-260.
4. Klueberg, O. "Cultural Factors in Intelligence Test Performance." *Journal of Negro Education*, III (1934), 478-483.
5. Lewinski, R. J. "Discriminative Value of the Sub-Tests of the Bellevue Verbal Scale in the Examination of Naval Recruits" *Journal of General Psychology*, XXXI (1944), 95-99.
6. Margaret, Ann and Wright, C. "Limitations in the Use of Intelligence Test Performance to Detect Mental Disturbance." *Journal of Applied Psychology*, XXV (1943), 387-398.
7. Price, J. "Negro-White Differences in General Intelligence." *Journal of Negro Education*, III (1934), 424-452.
8. Rabin, A. I. "A Short Form of the Wechsler-Bellevue Test." *Journal of Applied Psychology*, XXVII (1943), 320-324.

9. Reichard, Suzanne and Schafer, R. "The Clinical Significance of Scatter on the Bellevue Scale." *Bulletin of the Menninger Clinic*, VII (1943), 93-98.
 10. Rosenzweig, S., Bundas, L. E., Lumbry, K. and Davidson, Helen. "An Elementary Syllabus of Psychological Tests." *Journal of Psychology*, XVIII (1944), 9-40.
 11. Schafer, R. and Rappaport, D. "The Scatter in Diagnostic Intelligence Testing." *Character and Personality*, XII (1944), 275-284.
 12. Wechsler, D. *The Measurement of Adult Intelligence* (second ed.). Baltimore: Williams and Wilkins Co., 1941.
 13. Wechsler, D., Israel H. and Balinsky, B. "A Study of the Sub-Tests of the Bellevue Intelligence Scale in Borderline and Mental Defective Cases" *American Journal of Mental Deficiency*, XLV (1941), 555-558.
 14. (Anon) "A Selected Bibliography on the Physical and Mental Abilities of the American Negro." *Journal of Negro Education*, III (1934), 548-564.
-

MEASUREMENT ABSTRACTS*

Alper, Thelma G and Boring, Edwin G "Intelligence-Test Scores of Northern and Southern White and Negro Recruits in 1918" *Journal of Abnormal and Social Psychology*, XXXIX (1944), 471-474

Criticism of Benedict and Welfish's *The Races of Mankind*, which raised a controversy by presenting evidence to show that there was no relation between skin color and intelligence, is made on the grounds that their selection of data is open to censure. By use of an analysis of variance technique, it can be shown that "skin color as well as geography did affect the test scores of recruits in 1918." It would have been better, therefore, if Benedict and Welfish had given all of the data, and then gone on to argue that it is the Negro's educational disadvantage which handicaps him in such situations. *Lorraine Bouthilet.*

Bolanovich, D J "Selection of Female Engineering Trainees." *Journal of Educational Psychology*, XXXV (1944), 545-553

Eighty-six women were selected and trained in a ten-months' electronic engineering course. The data analyzed included test and rating scores, final grade-point averages (GPA) for the course, and termination records. The author found selection was based primarily on interviewer's over-all judgments of fitness. GPA had significant correlations with *American Council on Education Cooperative General Mathematics Test for High-School Students*, the *Wonderlic Personnel Test*, previous school grades, "fitness" rating, and "personality" rating. In comparisons between high and low achieving students and terminating students, the ACE mathematics test, the *Wonderlic Personnel Test*, and the *Kuder Preference Record*, computational key, showed significant differences. *E. C. Bell*

Bradway, Katherine P "IQ. Constancy on the Revised Stanford-Binet from the Pre-School to the Junior High School Level." *Journal of Genetic Psychology*, LXV (1944), 197-217

This reports a follow-up study of 138 children, comprising two groups between ages 2 and 6, who were examined on both Forms L and M of the *Revised Stanford-Binet Scale* during its standardization, and then retested 10 years later on Form L. Previous studies of IQ. constancy, involving initial tests at the pre-school level and retests at varying intervals, are cited for purposes of comparison and contrast with the author's findings. Correlations ranging from .58 to .67 for both groups and both forms indicate, in the author's judgment, a significant predictive value for the *Stanford-Binet* equalling, if not surpassing, other tests, and assure the importance in prognosis of the pre-school IQ. for the group and for the individual when accompanied by supplementary data. *Vernon S Tracht.*

Brown, Fred "An Experimental and Critical Study of the Intelligence of Negro and White Kindergarten Children." *Journal of Genetic Psychology*, LXV (1944), 161-175.

A group of 341 native white children of Minneapolis were compared on the *Stanford-Binet*, Form L, with 91 Negro children of the same city. The mean age for the white group was 69.51 months as compared with a mean age of 69.15 months for the Negroes. The mean I.Q.'s for the white and Negro groups were 107.06 and 100.70, respectively. A comparison of the intelligence of the two groups at various occupational levels reveals that the total Negro group resembles the white group at the semi-skilled and unskilled labor class. The results differ from

* Edited by Forrest A Kingsbury.

previous studies. The conclusion of the author is that the developmental construction of the Negroes is based upon cultural factors. *Betty Steele.*

Burt, Cyril. "Statistical Problems in the Evaluation of Army Tests." *Psychometrika*, IX (1944), 219-235.

The introduction of psychological tests for personnel selection in the British forces has given rise to several novel problems in statistical procedure. The solutions proposed are in the main extensions of devices already familiar in educational psychology. The more important are. (i) where the criterion yields a threefold classification only, a method of triserial correlation or of biserial correlation assuming point-distributions for the extremes; (ii) where the data on which validation has to be based are drawn from a selected sample, a simplified form of Pearson's equations to correct for selection; (iii) where the best line of demarcation has to be deduced from theoretical rather than practical considerations, a formula based on the principle of minimal discrepancy. (Courtesy *Psychometrika*)

Cattell, Raymond B. "'Parallel Proportional Profiles' and Other Principles for Determining the Choice of Factors by Rotation" *Psychometrika*, IX (1944), 267-283.

The choosing of a set of factors likely to correspond to the real psychological unitary traits in a situation usually reduces to finding a satisfactory rotation in a Thurstone centroid analysis. Seven principles, three of which are new, are described whereby rotation may be determined and/or judged. It is argued that the most fundamental is the principle of "parallel proportional profiles" or "simultaneous simple structure." A mathematical proof of the uniqueness of determination by this means is attempted and equations are suggested for discovering the unique position. (Courtesy *Psychometrika*.)

Goldfarb, William. "Adolescent Performance in the Wechsler-Bellevue Intelligence Scales and the Revised Stanford-Binet Examination, Form L." *Journal of Educational Psychology*, XXXV (1944), 503-507.

Scores of 60 adolescents living in foster homes and dependent for various periods of time, were correlated on the *Revised Stanford-Binet, Form L*, and the *Wechsler-Bellevue Scale*. The study confirmed the significant correlations between the I.Q. ratings on the two tests, but, unlike the findings of previous studies, the *Wechsler-Bellevue* I.Q. tended to be lower at all intelligence levels, especially so among children with *Wechsler-Bellevue* I.Q. of 110 or higher. This confirmed the author's practical experience that the *Wechsler-Bellevue Test* appears to be poor in discriminating the superior adolescents. He believes that, while test dispersion may partly explain the differences in I.Q. between the two tests, there is also a difference in the mental patterns of the groups studied. Therefore, he does not advocate a single regression formula derived from small samplings. *E. C. Bell*

Havighurst, Robert J. and Hilkevitch, Rhea R. "The Intelligence of Indian Children as Measured by a Performance Scale." *Journal of Abnormal and Social Psychology*, XXXIX (1944), 419-433.

In order to find out the ways in which the children of several Indian tribes varied from tribe to tribe and from community to community within a tribe, and also to compare their scores with those of white children, 670 Indian children ranging in age from 6 through 15 were tested on a shortened form of the *Grace Arthur Point Performance Scale*. The *Arthur Performance Scale* was used because previous studies have shown it to be relatively culture-free. It was found that Indian children did about as well as white children, and that tribal and community differences exist just as in various groups in a white population. There was some indication that children from tribes little influenced by white culture did not do so well on the test, but there was no evidence to support the statement that Indian children work more slowly than white children. It is concluded that with Indian children a performance test is a better instrument than a test requiring use of the English language. *Lorraine Bouthilet.*

Holzinger, Karl J. "A Simple Method of Factor Analysis." *Psychometrika*, IX (1944), 257-262

A simple method for extracting correlated factors simultaneously is described. The method is based on the idea that the centroid pattern coefficients for the sections of unit rank of the complete matrix may be interpreted as structure values for the entire matrix. Only the routine centroid average process is required. (Courtesy *Psychometrika*.)

Klugman, Samuel F. "Test Scores for Clerical Aptitude and Interests Before and After a Year of Schooling." *Journal of Genetic Psychology*, LXV (1944), 89-96

To determine whether test scores for clerical aptitude and interests, and the relationship between these two, remain the same after a year's schooling, 207 white, female, native-born students in commercial courses of a vocational high school were tested and, after 2 semesters' training, retested on appropriate portions of the *Strong Interest Blank* and the *Minnesota Clerical Aptitude Test*. A comparison of scores from the 30 oldest and a like number of the youngest indicated that the general improvement in scores noted for most subjects is probably due to schooling rather than maturation, since no reliable difference between means was found. Correlation between scores on the same tests one year apart revealed high relationship for clerical aptitude and substantial relationship for clerical interest. *Vernon S. Tracht*

Krugman, Morris. "Recent Developments in Clinical Psychology." *Journal of Consulting Psychology*, VIII (1944), 342-352.

Two general trends in clinical psychology during the war period are observed by the author. 1) Halt in research on new clinical techniques, and 2) Great advance in experimentation in and use of short procedures including group tests and screening methods. The Army's mental hygiene units are "child-guidance" clinics (for soldiers), emphasizing test patterning and diagnosis, factor analysis in evaluation of test batteries, and increased interest in projective techniques, especially the *Rorschach* and the *Thematic Apperception Test*. Abbreviated individual and group techniques are being developed for them. There is a corresponding loss of interest in personality questionnaire tests. Clinical psychologists are emphasizing diagnosis and neglecting psychotherapy. *E. C. Bell*

Richardson, Marion W. "The Interpretation of a Test Validity Coefficient in Terms of Increased Efficiency of a Selected Group of Personnel." *Psychometrika*, IX (1944), 245-248.

The predictive efficiency of a test used to select personnel is defined in terms of total effectiveness of the group thus selected, as compared with chance selection. The formula developed requires the use of an estimate of the ratio of average effectiveness of men selected to the average effectiveness of men not selected by the test. The predictive efficiency of the test varies directly with the magnitude of this ratio and also directly with the percentage rejected. (Courtesy *Psychometrika*.)

Sadowsky, Michael A. "Mathematical Analysis in Psychology of Education. Computation of Stimulation, Rapport, and Instructor's Driving Power." *Psychometrika*, IX (1944), 249-256.

Mathematical expressions are derived for such concepts as stimulation of student by instructor, student-instructor rapport, and driving power of instructor, in terms of the student's and the instructor's foci of attention, their strength of concentration, and the intensity of the presentation and of the reception of details of subject matter. Under the assumption of normal distribution, the mathematical methods of combination and integration yield conclusions on summary integral effects of interrelations within the educational team. The psychological interpretation of the mathematical results thus obtained conforms with common sense. The main emphasis of the article is the exposition of how the mathematical method of combination and integration can be used to estimate the resultant effect of various independent combined simple factors acting independently within the in-

dividuals forming the educational team. No claim is made as to the absolute truthfulness and reliability of the psychological postulates used at the beginning stage of the mathematical analysis. (Courtesy *Psychometrika*)

Spinelle, Leo and Nemzek, Claude L. "The Relationship of Personality Test Scores to School Marks and Intelligence Test Scores." *Journal of Social Psychology*, XX (1944), 289-294.

Results of a study undertaken to investigate the usefulness of the *Link Inventory of Interests and Activities* for prediction of success in school showed that, with junior-high-school girls, the measures yielded by Link's scale "do not possess direct value for educational guidance." It appeared from the fairly high correlation between intelligence quotients and school marks that the intelligence quotient could be used for group, but not individual, prediction of scholastic success; and that the *Link Inventory* should be considered as an objective questionnaire giving information to serve as a basis for discussion in personal interviews in a mental hygiene program. *Lorraine Bouthilet.*

Staff, Personnel Research Section, Classification and Replacement Branch, Adjutant General's Office. "The New Army Individual Test of General Mental Ability." *Psychological Bulletin*, XLI (1944), 532-538.

A new individual test of general learning ability was prepared in response to many requests from psychologists in the military services, especially those working in Special Training Units, Replacement Training Centers, and Army hospitals and convalescent centers. Seventeen verbal and non-verbal tests were tried out, the reliability estimated according to the Kuder-Richardson formula, and validation carried out with the *Army General Classification Test* as the criterion. Three verbal tests and three non-verbal were chosen on the basis not only of statistical considerations but also of several practical requirements making the test applicable for Army use. The test was standardized, and norms are given in terms of standard scores and Army grades. *Lorraine Bouthilet.*

Wallen, Richard. "Some Testing Needs in Military Clinical Psychology." *Psychological Bulletin*, XLI (1944), 539-542.

Tests developed in civilian life are sometimes not applicable to military needs, especially in the task of testing recruits. Most published tests are too long, too dependent on a high level of reading ability, and too much time is needed for scoring and interpretation. A test for recruits should have easily understandable directions, the performance required should be simple, and the reliability and validity should be based on appropriate norms. It is possible to construct such a test because the problem is primarily one of discrimination at only one end of the trait continuum—that is, of determining men who are not suitable for military service. Since the purpose of the test is to weed out the grossly atypical individuals, items to which a large proportion of the population respond in a given way are most useful. Promising results have been obtained in a few exploratory studies. *Lorraine Bouthilet.*

Wellman, Beth L. "Binet IQ Changes of Orphanage Children: A Re-Analysis." *Journal of Genetic Psychology*, LXV (1944), 239-263.

A pre-school and a control group of 47 and 44 children, respectively, were given the *Stanford-Binet* tests at the beginning and end of the project period which ranged from 77 to 972 days. The mean age for the pre-school group was 40.3 months as compared to 40.0 months for the control group. The mean I.Q. of the pre-school group was 86.9 while that of the control group was 83.5. The results reaffirm the original study, indicating that the pre-school child with regular attendance, and in residence more than a year, made significantly better progress in intelligence than the child of equal initial intelligence, and in residence for a similar period, who did not attend pre-school. *Betty Stacks.*

Wherry, Robert J. "Maximal Weighting of Qualitative Data." *Psychometrika*, IX (1944), 263-266.

A method whereby biographical or other questionnaire data of a purely qualitative nature may be used to predict success or failure on an independent criterion

is presented. The method is not new but the present least-squares derivation and the transformation equation for punched card coding were not available in the literature. The proper weights are found to be proportional to the per cent of passers in the various categories. The method is suggested as a suitable substitute for non-linear approaches in connection with purely quantitative data as well. The implications of reweighting in connection with multiple regression are discussed. The lavish use of degrees of freedom makes cross-validation extremely desirable (Courtesy *Psychometrika*)

Wherry, Robert J. and Gaylord, Richard H. "Factor Pattern of Test Items and Tests as a Function of the Correlation Coefficient Content, Difficulty, and Constant Error Factors." *Psychometrika*, IX (1944), 237-244.

A dilemma was created for factor analysts by Ferguson (*Psychometrika*, 1941, 6, 323-329) when he demonstrated that test items or sub-tests of varying difficulty will yield a correlation matrix of rank greater than 1, even though the material from which the items or sub-tests are drawn is homogeneous, although homogeneity of such material had been defined operationally by factor analysts as having a correlation matrix of rank 1. This dilemma has been resolved as a case of ambiguity, which lay in (1) failure to specify whether homogeneity was to apply to content, difficulty, or both, and (2) failure to state explicitly the kind of correlation to be used in obtaining the matrix. It is demonstrated that (1) if the material is homogeneous in both respects, the type of coefficient is immaterial, but (2) if content is homogeneous but difficulty is not, the homogeneity of the content can be demonstrated only by using the tetrachoric correlation coefficient in deriving the matrix; and that the use of the phi-coefficient (Pearsonian r) will disclose only the non-homogeneity of the difficulty and lead to a series of *constant error* factors as contrasted with *content factors*. Since varying difficulty of items (and possibly sub-tests) is desirable as well as practically unavoidable, it is recommended that all factor analysis problems be carried out with tetrachoric correlations. While no one would want to obtain the constant error factors by factor analysis (difficulty being more easily obtained by counting passes), their importance for test construction is pointed out. (Courtesy *Psychometrika*)

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME FIVE, NUMBER TWO, SUMMER

<i>Philosophy and Practice of Personnel Selection.</i> HERBERT A TOOPS	95
<i>The Counseling Program of the Veterans Administration.</i> CARLOS E. WARD and GWENDOLEN SCHNEIDLER	125
<i>Personality Traits Associated with Abilities. I. With Intel- ligence and Drawing Abilities.</i> RAYMOND B. CATTELL ...	131
<i>Factor Analysis of Occupational Aptitude Tests.</i> STAFF, Division of Occupational Analysis, War Manpower Com- mission	147
<i>The Interests of Forest Service Men</i> EDWARD K. STRONG, JR.	157
<i>A Group Testing Program for the Modern School</i> WARREN G. FINDLEY ...	173
<i>Replies of Psychologists to Several Questions on the Practical Value of Intelligence Tests.</i> ARTHUR KORNHAUSER ...	181
<i>Measurement Abstracts</i>	191

PRINTED IN THE UNITED STATES OF AMERICA
THE SCIENCE PRESS PRINTING COMPANY
LANCASTER, PENNSYLVANIA

PHILOSOPHY AND PRACTICE OF PERSONNEL SELECTION

HERBERT A. TOOPS
Ohio State University¹

By definition selection implies more candidates than jobs, a choosing of the most fit. In boom years and wars selection wanes; it waxes in depressions and peace. Following the war it will become important again.

It has become obvious that much of our material progress is due, on the one hand, to a very few expert people who are able to invent such things as B-29's, radar, dehydration, and penicillin; and equally, on the other hand, to a multitude of Joe's, Bill's and Sally's whose skill of hand, keenness of eye and sureness of touch, in small things, just as surely is an expertness of its own. Some kinds of people do each of these respective kinds of work better than others. Subdivide and specialize industry as much as you will and still there will be more work for each of these kinds of people to do besides all the more supplying work for a third class of experts, the managers, the Henry Ford's, the Henry J. Kaiser's, the J. F. Lincoln's, and others of lesser publicity and prominence. Expertness is important in all these realms.

Selection is both positive and negative. When looking for traits that are rare—in consequence of which we pay well for them—we wish to include as many as possible of the desired traits in one man; we look ideally for the one man of all men who most completely can fill the bill, the one man who includes in his make-up all the positive virtues. Of Tom, Dick, Joe, and Sally there are a myriad; hence they are paid chiefly for their time rather than for their pattern of abilities, and here we may seek only to *exclude* a certain few undesirable or nega-

¹ On leave with the National Roster of Scientific and Specialized Personnel.

tive traits such as dishonesty, unpunctuality, sickness, lethargy, and the like. We can afford—or so we often think—to hire them on an actuarial basis of “hire ten and fire two” since “not so much risk is involved in any individual mischoice” and “if they are not expert we can quickly and easily train them.”

We also rely on the fact that what one workman fails to produce another may make up for by greater diligence; the shortcoming of one thus is offset by the superiority of another. In the creative, or inventive, and the managerial realms, on the other hand, the weakness of the one—whether superior or subordinate—is less compensable, and in fact is more likely to result in a weakness of both. Yet even here the idea has potential merit: Could one pick “superiors” to work in pairs—or in the general case in teams—so that the strengths of the one overcome the weaknesses of the other, and vice versa; the too conservative tendencies of the one curb the too radical tendencies of the other; the inventive tendencies of the one stimulate the productivity of the other, till in the end, over the years, both come, like long-wedded couples are supposed to do, to resemble each other highly in all the good and virtuous traits? In fine, what persons should work with what persons; what roommates room with what roommates, what persons be friends and pals with whom, and who should marry whom? We should consider such questions in terms not only of the traits they each possess but also of the traits they may wish to develop. This, of course, is a discipline mainly for the future. Its statistical difficulties are mainly a difficulty of notation. We have, so far, a dearth of studies along such lines. It involves questions not only of comparing profiles of persons in their present or cross-section aspects, but also of considering them in respect to their productivity and their probable future trend.

About the course of growth curves we are not so certain as we once were. Thanks partly to war, we are not such dyed-in-the-wool hereditarians as formerly. In war—which incidentally has come to America with distressing regularity over the decades of her existence, itself owed to war—the second and third generations which fight the succeeding wars have

in each succeeding crisis little precedent to guide them. Hence, in these latter days, now that scientists are recognized more for what they are capable of doing than merely as so many additional units in the supply of cannon-fodder, more and more innovations are being tried out. We discover that illiterates, not suffering from illiterate intellects, can be made "literate" in a matter of weeks; that most of the color-blind can be made color-seeing; that those deficient in the ability to see in the dark can be taught to develop "cat's eyes"; and that plodding farmers' sons may become great heroes of the air force. Do these not shake our faith in predestination pronouncements of a generation ago?

We sometimes lose sight of the fact that man, with his superior cerebrum, is the most adaptive mechanism there is, extremely sensitive to the world about him, and particularly to that portion of his environment marked off as "the people with whom he works and lives." So true is this that for every man who stumbles or falls, we, as practical psychologists, look for the woman in the case; for every divorce we suspect the spouse; for every turnover we suspect the foreman; and for every corporal's failure we suspect his sergeant. In a very real sense *the traits of one's fixed-relation associates thus are one's traits*. Enrolling such other-person's-data in parallel columns of the data book after the "personal" data, one has the makings of a two-curves profile involving this social relationship; amenable to all the techniques to which any data may be put, and with some inherent niceties, such as, for example, entering data by pairs (the paired associates' respective scores in a given trait) into a multiple-ratio regression equation. Compensation of traits would be revealed, perhaps, by opposite signs of the two associates' scores in a given trait.

We are becoming more concerned, too, about what are traits. If they are not inherited, or not so much inherited, what then? Do they boil down to some physiological tendency such as the fund of usable energy which the individual possesses, paralleled in turn by some simple index of amount of absorbability of vegetative tissues, or some set of more complex reasons geared up with the glands or the like?

In addition, we have come, on the one hand, through the work of the clinicians, who concern themselves with human motivation and more recently with human reformability, and, on the other hand, through the work of the statisticians who have concerned themselves with those individual cases which in correlation plots destroy the validity of the predictions, to question whether our erstwhile conception of personality as the added sum of its parts—or of its weighted parts—is the correct one. Everyone knows that some traits are more important than others—for particular purposes, or at particular times, especially at crises. This, then, is not to question the usefulness to society—which as a whole always works on the actuarial basis—of the concept of the ordinary multiple regression equation. The statistician will never be content, however, with validities in the fifties and sixties and an occasional higher one. He knows that the difference between an ideal and unattainable unity and his (always!) inferior index is in part due to, or is associated with, individuals who do not fit his simple hypothesis.

Add traits, either in numbers or of varied kinds, or both, and still the divergent individuals are almost as divergent as before; and the common observation is that despite all the work the multiple correlation coefficient is singularly unaffected. Can this mean that our form of regression equation is at fault implying in turn a wrong conception of the matter? As biometrists we are little concerned with the man who took to drink and went all to pieces when his wife died, the child who suddenly turned truant, or the genius who sometimes emerges from the two-room sod house on the prairie. [There are many more of such habitations (environments) than of mansions.] In general, we leave those "situations" to poets. Ought not our conception to be enlarged so that these cases, as well as those of normal school-children, of all ages, who hitherto customarily have been the subjects of our studies, are all subsumed under the same formula? There is little merit in explaining them away by naming them cases of shattered, peculiar, abnormal, or emerging personality.

Are there not indeed crucial traits which unless operative

at least to a minimum degree render all others fruitless? Is it conceivable, for example, that any American without a certain mathematical background (which few Americans possess but which many conceivably might acquire) could read Einstein no matter how intelligent, persevering, or able as a reader he is? This essential lacking, the product is zero, or virtually zero, no matter how favorable the scores in the remainder of the "causal variables."

In the statistical treatment both the units of measurement—in their mathematical aspects—and the form of the "regression equation" will surely be involved. Little is written on the matter now, but much more deserves to be and will be written about it in the years to come, particularly if a few mathematically capable or promising and ambitious people can be recruited to the psychological profession.

Finally, to round off our background of the matter, is it not clear that the cross-section aspect of a trait which we get as the result of a test or inventory (qualification form, questionnaire or interview, for example) is only *evidence*, rather than fact, only a straw in the wind of how the individual growth curve blows—or grows? ✓

There are available but few growth studies of individuals in the several functions of physical growth, mental growth, and social growth. Those few suggest a functioning interdependence of such growths, predicated perhaps on a common fund of life-energy basic to the several growths resident within. They imply that the child is father to the man in more than years. It is useful then to look at the child if we would see the man, for what he is and yet may be. (That old dogs prefer to learn no new tricks is more of a truism than that old dogs can learn no new tricks.) And it is useful to keep track of him throughout life. Consequently a general growth study of the individual may be more revealing than any number and any quality of cross-section variables which ordinarily we may collect for selective purposes. Quasi-growth studies may be made from a consideration of the ages of the individual, computed for the dates of certain happenings, such as of his several job promotions, acquisitions of responsibility and title, and the like.

Present prognostic tests may be more so by name than by realization, particularly in the absence of long-range follow-up studies to convict us of our errors of "misappellation." The value of an "evidence" is measured not by its name but rather by its validity coefficient. Thus we see that the improvement of selection lies largely in the development of new techniques. Only a very few of its roots go back to the past. Our present practice accordingly may be found to be faulty, or be buoyed up unduly only by actuarial considerations which, partly or alone, save the day.

The last point is worth belaboring.

Two hundred men are let out of a modern plant, A, let us say, which is closing down on war production, on a Saturday. They have been trade-tested and it is known that 80% of them are "good workmen." The other 20% of "not so good" workmen are able to drive rivets, bolt-up nuts, and in general do any work for which they were hired, but with inferior speed and accuracy. In peacetime they would be "non-hirable" because too inefficient. Plant B is just starting up a new department. It advertises on Sunday for 100 men to show up at its employment office on Monday morning. At seven o'clock Monday morning let us assume that all of the recently discharged 200 are standing in line outside B's employment office door; also that in recent days the government has doubled its initial order with B, so that the word goes down to the employment office to "Hire them quickly so that we can get them to work. We can fire those who don't fit in. The government will pay the bill." So our obliging employment clerk counts the men in line, notes that there are 200, mentally calculates, since they want only 100, that "flipping a half dollar will do the job nicely and moreover will give every man an equal chance at a job." He makes a little speech to that effect and hires the men accordingly. And all are happy, even the men who get no job—since they had their chance—as though that were a good in itself—save the father whose infant daughter badly needs at once an operation which only a job can pay for. (This is statistics and probability, not humanitarianism or social security.) So in fifteen minutes the impartial half-

dollar responds to "Heads we hire this man; tails we don't" in its expected manner and soon one-half of them, 100 in number, are at work. Silly? Perhaps! But what is the worst job that could possibly be done by our half-dollar under the circumstances? The answer evidently is: To hire all 40 of the "not so good" men. (Twenty per cent of 200 men equals forty men.) That would yield a selection efficiency of 60 per cent. The half-dollar would do the job as badly as that "only once in a million times." What is the best it could do? Obviously, to hire 100 of the 160 "good men," and reject all 40 of the "not so good" as well as 60 of the "good," with a resulting index of efficiency of 100 per cent. This too is a very improbable occurrence. What is the most probable result? Evidently to hire 80 good men and 20 not so good ones, resulting in 80 per cent of efficiency of placement. (The same figure which the trade-tests revealed.) If classification experts, placement clerks, manning tables, job families, morale officers, and all the rest, can beat this figure, the excess—above the 80 per cent—can be credited to their efforts; and if a lesser figure is obtained, their value then is all on the negative side of the ledger!

Let us carry this matter a little further back in time. We note, then, that much selection has occurred before the 200 men started for the doors of our employment clerk early this morning. This may conveniently be said to consist of two types: self-selection and pre-selection.

The self-selection consists in such facts as that only men able to walk showed up at the office; only men able to read or possessed of relatives and friends who could read; that no coffee-tasters, goldbeaters, astronomers, mind readers, college professors or tight-rope walkers applied—indeed only people who wanted work, *this* work which they judge, rightly or wrongly, is like what they have done before or which they hope they may be able to qualify for.

The pre-selection consists in the effect which the ad, mainly perhaps, had on the potential applicants. If the ad says "gentiles only," few Jews will have the temerity to apply. If it says "white or colored," some colored who otherwise would

be hesitant may decide to apply. If it says "men only," what woman will apply? If it says "\$10,000 and expenses" what \$200-a-month clerk will apply? Indeed selection is benignly affected, all unknownst, by a myriad of such considerations which cost little or nothing of effort or thought—and which, in passing, may have just as little to do with the efficiency on the job of those hired, the lucky ones we would call them a mere four years ago, and perhaps still might do so with almost as much justice.

That the patient gets well without doctor is counted a miracle and that he gets well in spite of the doctor is never counted a blessing! If the doctor is called, all the world including the patient is satisfied that a great good has been achieved! But we—those of us appointed to note and to better such things—cannot be satisfied merely because our clients are contented! We find it necessary, useful, and right to look into the selection process with a critical eye. One fruitful way of doing this is to look into the implied statistics behind each of the current or possible modes of selection. It might be well, however, to delay that consideration for a little in order to ponder another matter or two and particularly one strictly statistical principle which grows out of the successive hurdles method of selection.² This method implies that only the passers of a given examination are allowed to take the subsequent qualifying examination; and evidently this test may also be a statistical one, a sieving process applied to the data of the several individuals on a common trait-profile recorded in their record cards; only those who qualify on the first trait being allowed to be considered for passing the next test, the next trait, the passing point of which comprises the next hurdle. In high-jumping contests we do not necessarily rank everyone exactly in the order of their jumping ability by ruling out those who do not (we did not say "cannot") jump the bamboo at three feet, a *difficulty* measure we say, forgetful of other facts such as that the motivation of severe competition often determines whether a record is broken. In "tests" of selection, not the difficulty, but rather the validity, of the test is the impor-

² Toops, Herbert A. "The Successive Hurdles Method." *The Personnel Journal* 11 (1932), 216-218

tant consideration. We get more or fewer jumpers merely by altering the height of the bar, by lowering or raising the passing point of the test. The all important consideration in selection is what kind of persons got over the bar—passed the test—and are allowed to take, or are subjected to, the next test. Accordingly if a trait of low validity is used for the first sieving any successive sieving of humans in selection, however fine, cannot undo the effect of having let most of the better individuals go into the discard. Still more concretely, the ends of all selection are at least two in number:

1. To assure a quality of those selected which shall be materially above that of those rejected.

2. To reduce a larger number of "applicants" to a smaller number of those recommended or hired.

The former objective, of course, is the important one. It follows then that in attaining the second end, that of reducing the number of candidates, it is important to keep the merit of the retainees as high as possible and to maintain it there as long as possible in the process of reducing the number of applicants. Briefly, it finally amounts to this: *that the ideal selection is achieved when the successive hurdles or tests are applied in descending order of their validity coefficients and that the application of additional tests is stopped arbitrarily whenever the number of applicants decreases beyond a given minimum.*³ If the tests have no validity at all, as is true of our half-dollar above, then the order of application of the hurdles is immaterial. All that the tests accomplish in this latter case is:

1. A considerable reduction in numbers by reason of the application of the n hurdles.⁴

³ If, regardless of the number of persons who pass all the hurdles, it is decided that all the hurdles (or tests each with their several passing points) shall be applied, willy nilly, the order of application of the tests makes no difference; the same individuals will respond to the total process in every permutation of order of application of the selective variables. We assume that the passers of two tests will be more able than the passers of one only; but this expectation follows, the first test having positive validity, only if the second test has a validity higher than zero; and it can actually destroy some of the validity obtained if the second is of negative validity. The validity of two can be less than the validity of the best one alone also when in weighting methods of selection we "overweight" the poorer test beyond its multiple regression β -weight. It is by no means axiomatic that "two tests are always better than one."

⁴ Five such successive hurdles, each eliminating 50 per cent of the previous retainees will reduce a field of 3200 applicants to 100. Two more, or seven in all, will reduce it to 25.

2. A building up, in the minds of the applicants, of a belief in the fairness of the tests. [A feeling which unfortunately can be, and very probably will be, almost as great for low validity, or zero validity (chance) tests, as for highly valid ones.]

So long as they were believed in, for example, tests of memorization of Confucius were "good tests" in China, but any statistician could tell you that the character of the civil servants of China under such a system could be no better than that of the average person able to master this first hurdle to political preferment. A good rote memory obviously is a valuable asset of a statistical coder but surely is not all important for a high civil officer—or at least most civil officers—of the state. This settles at one stroke any contention that a uniform pattern of selective traits—at least of those presently available—could be equally valid for all purposes, for all occupations for instance. Traits are important—or so we now consider—in terms of their weights in a multiple regression equation. Alter our conception of the regression equation and some new measure of the relative importance of the traits results. At present the traits employed by sociologists, census-takers, and newspapers to typify the social-composition of the individuals of a social body may have zero or even negative validity—depending on the extent of their correlation with an independent criterion of ability in the occupation or undertaking in question—and yet may serve just as well as any more valid set of traits to quickly reduce the number of applicants to a number feasible for "more detailed examination (usually an interview) and consideration before appointment."

Where selection of equals is at stake and competition for jobs prevails, clearly justice requires that each person be given an equal chance of being selected. Of course this principle does not apply to unequals, for the better man, other things equal, should always be chosen. And if things are not equal but are almost equal, then there surely is some social gain if the renouncement can be voluntary. Not every teacher, for example—who incidentally happens to have a husband—can see any justice in giving up her job to an inferior, equally good, or even slightly better teacher who happens to be both husband-

less and jobless. No obvious improvement—and possibly some loss—in the condition of the taught is foreseen as the result of such replacements. We are only in process of establishing the morality that one's right to a job and to advancement is independent of his social and economic status.⁵

The severity of the critical scores employed in the successive tests has a bearing not only on the amount and the quickness of elimination but also upon the validity of the sortings. To eliminate no one by a highly valid test is to eliminate entirely its validity as a selective agent, to cancel it out as if it had never been given. We are unable to state this effect in other than the very general statement⁶ that the more valid tests should have the more rigorous critical scores if there is any variation of standard in the several hurdles.

There was an intimation above that under some circumstances there was merit in an actuarial hiring of all who apply and in letting the test of the job decide who should be retained. On a cost-plus economy this is always feasible particularly if:

1. The percentage of applicants who are above the minimal "acceptable point" of competence is high. (In our example above, if only 50 per cent of our applicants instead of 80 per cent were "competent" the impartial half-dollar *could* give us no satisfactory selectees at all. With the number of competents less than 50 per cent, the probability of such an occurrence is higher and higher as the percentage index descends.)

2. There is some highly efficient method of quickly weeding out the incompetents on the basis of their performance subsequent to selection.

These considerations yield us our first potential or actual selection method.

The Test of The Job

If we measure day by day the rate of learning (or "progress") of beginners, say, by means of adequate work records,

⁵ The war has given us a new morality about the matter of one firm or one service gutting the market for good men, merely because it happens to be first, or have the most money, and the like

⁶ An important initial attack on this problem has recently been made by Richardson, M. W. "The Interpretation of a Test Validity Coefficient in Terms of Increased Efficiency of a Selected Group of Personnel" *Psychometrika*, IX (1944), 245-248.

both a prognostic test and a trade test may be unnecessary. Relative indices of progress on the job may then be all the test we need. This is true because there is some warrant for believing that differential capacities yield individual progress curves, with distinctly different rates of growth, which differentiate at an early date. The rank orders of the growth curves of the apprentices at an early date thus approximate the final orders of competence. Our technical problem then is to shorten the tryout period as much as possible.

Let us assume that all of the new workers in a certain factory are put through a sequence of jobs, first on the drill press, then on the shaper, then on the planer, etc. It is clear, then, that the number of hours that it takes candidate A to complete a standard drill press assignment as compared with the hours required by candidate B is some indication of which of the two is presently more competent thereon; perhaps some slight indication, also, of which of the two gives more promise—if both were equally “untaught” at the beginning—of future usefulness in that particular drill press department; that is to say, possesses more “drill press aptitude.” At the end of the ensuing shaper operations, which it is assumed follow in all cases upon the drill press operations, the *cumulative* number of hours required for completing both assignments gives a better measure of the “all-around” mechanical capability of the candidates than the previously mentioned shorter “test” dealing with only a single (somewhat more specialized) ability. Accordingly, the greater the number of pertinent experiences included in the testing the more is revealed the “all-around” mechanical ingenuity, or general mechanical ability or adaptability or learning-power, of the (apprentice) learners. The question now becomes a statistical one: “What is the earliest period at which the cumulated competence scores can be made to correlate to at least a minimum limit, say .90, with ultimate competence?” The end sought is to minimize the time, work and money necessary to make a valid decision of whom to keep (because worthy, probably, of promotion) and whom to discharge or transfer (because of inadequate aptitude for the work). The order of presentation of the work experiences, drill press-shaper-

planer must have been fixed upon, of course, and be uniform for a particular entering group of apprentices, but may be altered on subsequent groups to place the most valid shops (those correlating highest with the sum of them all) into the earliest positions in the try-out experiences, as is necessitated by the successive hurdles method. On our initial experiment all must complete all assignments to arrive at an over-all criterion score for each. Employing the L-technique one now may easily ascertain how highly the first shop correlates with the sum of the success scores involving them all; how high the score of success in the first two shops combined correlates with the total success; how high the combination of the first three, and so on. The correlation will become high at an early stage, particularly after the shops once have been arranged in an order of decreasing correlation of the several shops with the total success on all the shops.

An alternative, and preferable, method is to place in first or accepted position in our upbuilding composite job that task which correlates best with the sum of them all; then, in second place, that one of the remaining which raises the correlation most; then that of the remaining tasks which raises the correlation next most and so on. The eventual order is "that order which maximizes the prediction of the criterion with a minimum of tests." We have here a choice of utilizing either the multiple-ratio method⁷ or the L-technique, the latter being preferable wherever simplicity is important. The former does not require the computation of all the inter-correlation coefficients, which is an advantage if the number of elements to be combined is large. If all the intercorrelations are available then the Wherry-Doolittle technique is appropriate and has merit for the purpose of selecting a minimum of "shops" to correlate maximally with their total.⁸ This simultaneously minimizes the time, and optimizes the order of presentation, of the accepted shop experiences as a test of machine-shop aptitude.

⁷ Toops, Herbert A. "The L-Technique" *Psychometrika*, VI (1941), 249-266. See also, Toops, Herbert A. "A Self-checking Technique for Shortening a Test." *O C A Bull.* No. 92, Jan. 30, 1934, 2027-2040

⁸ Stead, Shartle, et al. *Occupational Counseling Techniques*. New York: American Book Co., 1940, pp. 245-250

In all such methods some regression of the validity coefficients of the selected scale, on subsequent tryouts, is to be expected. The point at which the thus obtained correlation, by whatever method, first reaches .90 or .95 say, will mark off both what shops it may be necessary to give to all candidates and also the desirable order of their presentation. The individual's score on the accepted composite will reveal substantially whether a given individual does or does not possess sufficient "aptitude" to be allowed to continue. Where there are necessary sequences in learning skills—of which there are few in either school or industry—the statistically dictated optimal order of presentation cannot of course be employed

Let us suppose that as a result of an adequate experiment in which a number of beginners have carefully been watched through to ultimate competence, it has been decided that the number of hours needed to complete the first five (thus reordered) operations is a highly valid measure for prognosticating ultimate and "all-around" success. It is clear, then, that if we build five successive sets of norms of cumulative progress as of the ends of these five several operations the "normated" individual performances obtained from the job itself on the five successive occasions of completion of an additional project are progressively more and more indicative of the ultimate worth or lack of merit of the individual apprentice of concern. For solution of our statistical problem no test, then, other than the test of the job itself, is necessary. Then if industrial agencies (as of wartime) demand a decision which may be made quickly with only fair accuracy it may be made after the first operation, with somewhat better validity after the second and so on, the validity improving with the length of the tryout.⁹ By the steps outlined we have discovered the minimal number of operations (involving their identity and sequence) which, with the least possible wastage of supervisory time and overhead cost, enables us to arrange the candidates

⁹ As in sermons, where it is said that most souls are saved in the first fifteen minutes, so with tests. Tests or hurdles beyond the first highly valid one add mostly to the reliability of the scale—and so to the "justness" of the exclusions, when considered from the view of the excludee or potential excludee—but add little to the validity of the scale.

substantially in the order of their relative excellence—centile ranks, perhaps—for retention (and promotion). The above applies when even the poorest man turns out a salable product.

But the problem is not quite so simple as the above statements would imply. "A man is worth the selling price of what he adds to the product less a reasonable profit" is an axiom in economics perhaps but vocational psychology is not quite so naive as that. Observation shows that the worth of a man is measured adequately neither by the wage paid him nor by the quantity of work he turns out. Both are only *symptoms* of his worth or value. The slow-learning but ultimately fairly capable employee may be more of a "success," from the employer's viewpoint, than the quick-learning person who is the delight of the teacher or foreman (the typical factory teacher). This statement follows if the employer values highly traits other than productivity, matters such as punctuality, dependability, or gentility. No series of measures can ever completely describe the man. The individual is what he is in the measures employed; he is something else when more or other kinds of measures or tests or aspects of worth are applied. The employer then may indulge his *preferences* (or prejudices) between such "types" as a "slow but accurate" individual and a "rapid but somewhat inaccurate individual." Concretely, let us assume that the first of two persons rates 40 for speed and 60 for accuracy. If the arithmetical units are comparable, so that they may legitimately be averaged, the first individual then averages out to 50. Let us suppose that a second individual rates 60 in speed and 40 in accuracy. His average also is 50. But the two individuals, thus seeming alike, each having an average of 50, are quite different in the employer's eyes. The first will not get so much work done, but he has the merit that he will not destroy so much raw material; the quality of work produced will be better; and he probably will not require so much supervision and correction of his work. Such considerations are "values" to most employers. Inquiry on the part of the U. S. Civil Service Commission, for instance, has revealed that business men are quite content with a lower typing speed in a secretary than that set as the required gradua-

tion speed by a majority of stenographic schools provided only that they can acquire people who do not make so many mistakes (even if they don't get so much done) which need to be corrected. In line with this, employers rate highly the ability to spell (to correct their boss' mistakes).

The above homely example indicates that success is not a "unitary" concept, it cannot be measured in one dimension but can be only more or less adequately portrayed by a profile, a series of scores in a number of variables, measuring ideally different "dimensions" of the applicant. Without tests, can we still measure the individual for selection in terms of a profile, to the end that the profiles of different candidates may be compared and a choice be made? By making the permissible assumption that in some occupations at least a man's past portends his future—which is a safe assumption in fields such as teaching, management, and the professions generally, wherever occupational growth requires considerable time and is continuous—this is readily done by a technique previously published by the writer,¹⁰ and may be briefly summarized as to its principles or elements:

The Test of Accumulated Evidences

1. An extensive qualifications blank and specific recommendations blanks received from former employers or teachers provide a great deal of concrete evidence, from various angles, as to an individual's success to date in his occupation and his adjudged promise therein.

2. By the aid of official evaluators of such evidence, who function as a standing jury in that capacity year after year, the lines of evidence are telescoped into quantitative scores, of some objectivity and validity, recorded as some fourteen (present practice) aspects of ability deemed important for the job in question: intelligence, research ability, scholarship in the specialty, general scholarship, social intelligence, managerial and executive ability, special skills useful in the occupation and the like.

¹⁰ Toops, Herbert A. "The Selection of Graduate Assistants" *The Personnel Journal*, VI (1928), 457-472

3. The resulting scores, when plotted, become the candidate's profile just as truly as if one had test scores, social composition items, standard scores on a standardized progress schedule, and the like.¹¹ The subsequent evaluation may now be by any of the general methods given below. In the article referred to, the subsequent evaluation was by the summation-of-traits-scores method, which may or may not be best for one's purpose, as noted hereinafter.

Given, then, a profile of the individual, assuming all necessary corresponding statistical implications, there are a good baker's half-dozen different ways of selecting talent, that is, of reducing the number, while increasing (as an expectation) the competence, of the retainees. All of these in common assume that the pertinent traits (or at least the traits on which he selection is to be made) have been observed, preferably objectively measured, on each of the candidates and that the scores necessitated correspondingly are available. ~

1. *The summation-of-traits-scores method*

1.1 *Where the scores are added at gross-score weights of 1.*

This is the method customarily used in most school examinations where the sum total of the person's merit is simply the arithmetical sum of all of the scores on the several sub-parts (sub-tests and items). This implies an addability of scores which seldom obtains.

This method, contrary to popular belief, does not weight the sub-parts equally, but instead weights each item, and each test, proportional to its standard deviation. In many cases this may not be such a bad assumption, however, since the test which can produce the greatest spread of scores in general is the most valid.

1.2 *When the several scores of a given person on several traits or sub-traits of the total examination in succession are multiplied by a fixed series of arbitrary individual test gross-score weights.* This is the "weighting" method which is popular with civil service gen-

¹¹ Toops, Herbert A. "The Criterion" *Educational and Psychological Measurement*, IV (1944), 271-297

erally It is likely that civil service frequently has greatly erred in the weight which it felt it was attaching to the scores on the several sub-parts in question.¹² It is a safe guess that, the standard deviations abetting, many a man has been selected by civil service mainly on account of his handwriting or some other comparatively unimportant trait. Such "errors" can be avoided by changing the original scores to ranks, standard scores, *T*-scores, or other relative measures in which the standard deviation of the different traits becomes a constant. Standard scores have a standard deviation of 1 while rank scores have a standard deviation of $\sqrt{\frac{N-1}{12}}$; *T*-scores a σ of 10, and so on, irrespective of the variable measured.

The weights may be ascertained by the bids procedure of the ensuing footnote¹³ which standardizes the process of ascribing the weights and replaces the presumably inferior judgment of a single person by the possibly superior judgment of a group of "experts."

The weights may be obtained mathematically, a criterion being available, and thus have the superiority ascribable to least squares technique. It will be recalled, however, that a unitary criterion score may come from vastly different profiles of "success variables."

The most popular of selective methods fails then to predict well even criteria whose similar mode of compiling undoubtedly weights the scales at least slightly in the direction of high, rather than low, validity. It is perhaps the simplest concep-

¹² If β_i be the importance actually assigned to standard scores, then $W_i = \beta_i / \sigma_i$, where W_i is the proper gross score weight to be employed in order that the trait shall be weighted with a true relative weight or importance of β_i . Conversely, if a gross score weight, W_i , be arbitrarily assigned to Test 1, the true relative weight or importance thus assigned is not W_i , but $\beta_i = W_i \sigma_i$. Clearly, β_1 and β_2 are proportional to W_1 and W_2 , or conversely, only when $\sigma_1 = \sigma_2$, or in general, only when the σ 's of the tests (of sub-parts) are equal. Standard scores, ranks, and *T*-scores among others, have this property.

Clearly also one may vary the length or difficulty of the examination in order to achieve approximately this end, a point of finesse worth further exploitation. If a test is lengthened to n times its present length its new and enlarged standard deviation is $\sigma_n = \sigma \sqrt{n+n} \frac{(n-1)\bar{r}_{11}}{n}$ where \bar{r}_{11} is its reliability coefficient

¹³ Toops, Herbert A. "The Selection of Graduate Assistants" op cit

tion of the organization of traits, merely the sum of the weighted scores, and is suspect for that reason alone—as is the I.Q.; for what else, pray, could one do in the latter situation but take a ratio! Man prefers simple laws; Nature, perhaps!

2. *The successive hurdles method*

This, referred to above as a principle, is the method which is popular with civil service in times when the supply of candidates greatly outruns the possible demand for placements. An examination is given in common to the entire list of candidates who have qualified on certain preliminary *minimum specifications* (health, freedom from arrest, etc.). A critical score is established above which all of those who “pass” the first examination are certified as entitled to take a second examination, while those below are denied further opportunity to qualify. Successive examinations eventually whittle down a large list of candidates to a few who are certified for appointment. The author¹⁴ has shown that, when using this method, it is highly important that the tests be administered in the order of the most valid test first, the next most valid second, and so on down to the least valid last. The least valid and last test may then be practically a chance test (It should not have a negative validity coefficient, of course, for that would reduce, instead of improve, the mean talent of the retainees.) Accordingly, if chance is to have a hand in the selecting of candidates who eventually are offered positions, clearly chance should operate only on a group of people who as a result of preceding screenings have a very high degree of capability. This condition easily is procurable by the sequence recommended. Clearly it were better that all the tests should be equally valid (and ideally intercorrelate zero), but since this is impossible the principle stated is obviously the correct one. From the viewpoint of the function performed it is clear that an individual always is thrown out by a single test, which in the nature of all tests has at least a degree of unreliability and a validity which may or may not be very pertinent, i.e., valid, to the particular job

¹⁴ Toops, Herbert A. “Sifting Civil Service Applicants by the Successive Hurdles Method” *The Personnel Journal*, 11 (1932), 216-219

which a given candidate is to fill.¹⁵ This method had its origin during the depression when thousands applied for positions for which scores or at least only hundreds had applied before. It is the product of desperation! It is costly; and at best is not the method which one of free choice, cost being no consideration, would choose.

For many a candidate his destiny is decided by only one or two traits (the earliest tested) of his entire profile. The average competence of 100 candidates chosen by such methods—if the selective traits are valid—of course is high¹⁶ when the tests are administered as above indicated, namely in a decreasing order of validity, and with the more valid tests producing the greater (and earlier) elimination of the applicants. This is the method employed by the Westinghouse Science Talent Search. Functionally, this method has much in common with the precise profile method below (q.v.). It has the merit that by setting a more severe standard the testing can be terminated briefly, thus minimizing its cost.

3. *The precise profile method*

In this method a certain number of traits are assumed to be highly (and equally) important to the extent that if an individual does not have precisely the skill-pattern, or profile, established as important (possibly by preferences of the prospective employer) he shall not be considered. The traits of a large number of individuals having been punched into Hollerith cards or Findex Cards, say, the subsequent sorting process yields all those persons who exactly fit the "selection profile." If too many candidates result, another trait, of little validity perhaps, may be added to the pattern to reduce still further the number of cases; or the subsequent selection may be subjective, based on "a detailed examination of the applicant's entire dossier." (This implies a positive validity co-

¹⁵ Obviously the larger the roster and the fewer the referrals the greater the competence of the actual referrals, if the given selective method has positive validity.

¹⁶ The normal expectation in lengthening a test is that this will increase greatly its reliability and increase slightly its validity. One of the "easy" ways to improve the validity of a test is to *prune* it by eliminating the least valuable items. Tests of 100 or more items often will be found to have items of negative validity in them.

efficient of the examiners' judgments.) Only those who qualify on *all* the traits are acceptable. The machines are blind; they know no "humanitarian considerations." The method, as outlined, presupposes that all the traits are applied without fail to the candidates; hence the order of sorting of the traits (if one was employed) can have no influence on the competence of the selectees. The method presupposes that the needed numbers of candidates have been measured on all traits, and the task is that of picking the "most likely" candidate for placement.

If Findex cards, and quantitative variables, or qualitative ones possessing intrinsic quantitative characteristics are employed the chief fault of the method may be somewhat mitigated by slotting¹⁷ on the "or more" basis in order to pick out, automatically, candidates who have *at least* the minimum standard, *or more*, on each trait. Thus, any person who is 9 in a trait is also slotted 8, 7, 6, 5, 4, 3, 2, 1, and 0; while any person who is 10 is also slotted 9, 8, 7, 6, 5, 4, 3, 2, 1, and 0. If, then, one sets, say, 7 as a minimum in a selective trait, all persons who are 8 or above and 7 or above also will respond to the selector rods of the mechanism, as well as those persons whose score is precisely 7. In the above case, if one sets 10 as his minimally acceptable score, only the second candidate will respond to the rodding; but if he sets 9, or 8, or 7, say, both candidates' cards will respond because both are equal to or are greater than these several limits, respectively. If not enough candidates show up, one may then reduce his minimum to 6, say, and find additional cases who are 6 on the test in question to "fill one's quota," on the assumption that one wants to pass some definite per cent of all the applicants on this one trait.

If the entire pattern of traits is specified at the outset and is applied to all the cases and this results in turning up several candidates for each one to be hired the final selection may then be made on the basis of "personality" traits or even on other considerations having zero (but not negative) validity. If too

¹⁷ If Hollerith cards are employed, the same end may be achieved by sorting the several traits *in succession*, picking up for subsequent sorting not only all those "passing" but also those of higher standing in the trait in question.

many candidates show up, one may raise his minimum requirement in the trait or traits in which it is known or is judged that such raising is likely to be most profitable (namely the trait or traits of greatest validity.)

Some of the traits may be expressed as *unvarying standards*; that is to say the particular category in the trait in question is the only one acceptable. Thus, one may insist on having a male teacher, females being totally unacceptable; or an unmarried teacher; or a Protestant teacher, or any compound characterization of these,¹⁸ e.g., a married-man, or a married-man-who-is-also-a-Protestant. With respect to traits which are strictly quantitative, such as intelligence, scholastic average, height, weight, etc., usually anyone *above* a certain minimum is deemed acceptable. Sometimes, however, an upper limit is established as well. That is to say, policemen must be at least 6 feet, 0 inches tall and not more than 6 feet, 4 inches. (This would imply the existence of a curvilinear relationship of the traits upon the independent measure (criterion measure) of job success, for which selection is being carried on.) If Hollerith equipment is employed similar conditions may be made to prevail either by multiple-punching one-column traits in similar fashion (this precluding tabulation) or, better, by the selection of particular packs of cards for *successive* sortings. The multiple-sorting head device is essentially the exact profile method.

This "or more"-variation of the method resembles the popular method used by housewives in picking maids, by police captains and army recruiting sergeants in picking recruits, by employers in hiring workmen, and the like. It is frequently used by placement bureaus to place individuals, the card being removed from the "live file" just as soon as a given person is placed. It is the method employed by the National Roster of Scientific and Specialized Personnel, with the variation that an already employed person, if an "essential specialist," may be referred to a prospective employer whose priorities warrant.

¹⁸ Only by an adequate follow-up, *when all are* hired (as in wartime) of each such compound type can we ascertain the relative merit of such qualitative qualifications. It is not impossible that some of these are on the negative side of the ledger. Humanity undoubtedly pays through the nose for its hiring prejudices.

Clearly the method presupposes a great supply of persons for a limited number of placements. Its essential difference from the successive hurdles method if the or-more punching or selection is employed, is a distinction without a difference. The fact that all the possible candidates have had all the "tests" (e g, have answered all the questions of all application blanks and supplied all required credentials) and that the entire pattern is employed makes not a whit of difference. A too-low score on only *one* trait eliminates him sooner or later, but just as surely. If the result of the compound dragnet (applying all the hurdles at once instead of in succession) is unsatisfactory, the test profile may be changed until a "likely" candidate comes to hand.

Its chief merit is its potential flexibility. The housewife, so inclined, for example, may interview all who answer her ad and if many respond raise her pattern of requirements to "get more for her money."

When the traits are set by employers ignorant of trait implications the result obviously can become little more than a means of securing a random selection among the available candidates and of indulging the whims and prejudices of the requisitioner. The best to be hoped for is that any such pattern in general has "some pertinence" to the jobs in question.

Its great fault is that the candidate who is generally very superior but lower than the minimum in only one trait—who is 100 centile, say, in all traits but one in which he is, say only one centile below the established minimum—will fail to be selected. And since the selection is blind—i.e., the operator sees only the cards of those who respond to the selection, and is blissfully unaware of the "all but" cases—the conspicuous failures of the method—its shortcomings are never noted.¹⁰ The Findex is limited to selection where only such numbers of cards (individuals) are involved as can be manipulated for hand sorting. Larger numbers could be handled by having

¹⁰ This presumably might not be the case with Keysort or Speedsort cards since the "almost" cards would have a single solid card across a continuous channel of holes if the card in question chances to be in the midst of the "all but one operation to be performed" cards.

the cards filed in pre-sorted compound-breakdowns, which to that extent may defeat its own end.

This might be called the fastidious method of selection. It reminds us of the diner who to quench his thirst must have wine; and not only wine but Chablis; and not only Chablis but Chablis of a particular vintage, 1927, say; and who moreover is willing to pay any price, including that of not drinking, in order to appease his whim! Unless the available supply of applicants is large the precise pattern may locate no one. This, however, oftentimes is the consideration which makes the method appeal to the politician who would nullify civil service; since with a not too large roster an employer given that option very easily may set up a pattern of traits that none but the intended recipient of the proffered position will be able to meet.

In practical application the practice is to establish *by judgment* an ideal pattern; and then to investigate the available supply of applicants by the Findex or by the more recently perfected Multiple Pattern Sorting device of the Hollerith machine, an attachment for the sorter which will sort out all cards of any given pattern not exceeding ten contiguous columns of the eighty-column Hollerith card. Then if the initial sort yields no applicant, or not enough applicants, the pattern may be altered to a less ambitious one, whereupon more applicants will come to hand. In theory the "less important" qualifications are dropped, but who, in the absence of appropriate regression equations can say which are the "less important" qualifications!

Where prospective employers set the patterns, if they could be induced to establish *an order of decreasing desirability* for sorting the cards, and if the number of desired referrals could be always specified also, or else be predeterminable by formula *from the number* of workers requisitioned, then the sorter would have a freedom which permits of lowering or raising the standard of competence in order best to fill the requisition. Even here "stereotyping" of orders may reduce greatly the chances of referral of a man who is substantially good but not in exact agreement with the stereotype. The short man who on other scores would be a superlative policeman is out of luck

so long as more than enough 6-footers respond to that universal requirement! And this despite the fact that modern automats are supposed, for some purposes, to make all statures even!

4. *The minimum divergence from the desired profile method*

Having an "ideal profile" in mind the candidate's profile of test scores can be matched against this profile and the candidate then can be hired by reason of an index of his agreement or lack of agreement with the ideal profile in question, say "that of the job." Probably one of the very best measures of the extent of such agreement is Sagebeer's index.²⁰ This index has a value 0 when the profile of the individual exactly matches the profile of the individual for which the selector is looking, while larger values of the index indicate greater discrepancies between the candidate's profile and the ideal. The method obviously assumes that the traits are *measured* scores and are *comparable*. Possibly the scores should be standard scores, thus making the assumption that equal standard scores in two or more traits are equal. The method will be most meaningful, perhaps, in large groups of applicants, all tested on quantitatively scored tests which are highly reliable and as nearly unique as possible. Possibly it applies better where aptitude rather than achievement is the basis of the selection. *The ideal of occupational selection, as of any other type for that matter, is that all applicants, not merely those who happen to be labeled by the particular magic name of an occupation, for example, should be scanned for selection.* The above index undoubtedly would be too laborious for this purpose without some machine method of solving the formula. This method, employing the index, has possibly never been formally used or used very much because of its newness. The method presup-

²⁰ Sagebeer's index is:

$$I = \sqrt{(X_{1I} - X_1)^2 + (X_{2I} - X_2)^2 + \dots}$$

where

X_{1I} is the ideal profile score in trait 1, that established authoritatively by job analysis.

X_1 is the candidate's score in trait 1

The correlation coefficient between the scores of the candidate's profile and the ideal profile would be an alternative index for the purpose at hand

poses that every person has been compared with the ideal profile and a measure of his disparity therefrom has been obtained. Assuming acceptable *comparable units* for the several tests, this is possibly the most ideal of all methods, for general selective purposes. It tends to prevent an individual who is very lopsided from being chosen and yet allows deviations from the range of ideal scores to be considered—although at a (slight but appropriate) disadvantage.

5. *The predominant or outstanding merit method*

The essence of this method consists of the following operations:

- 5.1 On a sheet of cross-section paper, in Column 1 are entered the names of the several candidates.
- 5.2 The headings of the several columns are labeled with the various traits on which the candidates are measured.
- 5.3 Into the compartment is written the most significant statement (preferably quantitative) of the candidate's standing in the trait in question. This is done for all traits and all persons in turn. This then is the primary data table, or the Hollerith tabulator reproduction thereof.
- 5.4 Considering now trait 1 only, one may encircle in column 2 those statements which represent the *most meritorious* degrees of the trait desired in the candidate. [In general these will be the largest scores; in some cases, such as errors or time, they will be the smallest scores; and in case of curvilinear relationship they will be the scores that indicate greatest fitness for the job, that is to say, the scores closest to the apex of the parabola or curvilinear relationship line between the trait in question (X) and probable job success (Y).] Possibly the ten per cent or so of highest scores are so encircled. The number so to be encircled will depend both on the number of candidates and on the number of traits.
- 5.5 The same now is done for the remainder of the traits.

- 5.6 One may now either count and record the number of circles for each person (that is, the number of circles *per row*) or, as an alternative, may count up, for each person, the sum of the *weights* of the traits appropriate to the columns in which a given person possesses circles.²¹

The end achieved by this method is to give unusual weight in selection to those persons who possess *unusual excellence in more than one trait*. In other words, it is rather the antithesis of the method in which we set up minimal test scores and are satisfied with any one who, so far as the single trait is concerned, is at least above that point. By this method we assure ourselves that the appointee surely will have several strengths—if any candidates possessing such are available—although he also *may* have some very fundamental weaknesses; that is to say, he may be lopsided, but nevertheless he will be at least a “near-genius” in some respects. The method presupposes a great scarcity of the kind of talent (genius) desired and implies a desire to locate the “best” that is available.

This method implicitly, rather than formally, is assumed in case one goes out “combing the world” for unusual talent of any kind. Unusual talent is so scarce and withal comes so dearly that if one can get a near-genius in several traits, one generally is quite willing to overlook a number of minor or even a few pretty serious faults or defects. The adaptability possible in a *specialized industry* often makes it particularly feasible to place an individual where his defects, from the viewpoint of the job at which he has to work, will be practically no handicap at all.

One may vary this technique by encircling instead all scores of defects which one *wants on all accounts to avoid* and then consider for appointment him who has the fewest defects. This, presumably, formally approximates the procedure ordinarily used in choosing diplomats and other public relations

²¹ The weights would be determined by the bids-procedure, presumably. If the β 's appear on a T-square blade, with the data table tacked in alignment to a drawing board, the β 's of the blade opposite the encirclements only are added. The T-square is a convenient device for bearing the marginal β 's into the body of the table.

appointees where freedom from public offense is or may be a greater virtue than positive capability.

Still another variation would be to encircle the strengths and then to box-in all weaknesses, the score for a person being the number of his circles minus the number of his rectangles.²² Such a method would require a larger number of registrants than either of the last two methods to produce the same number of selectees. It has the advantage that those selected have high basic strengths and at most only a very few fundamental weaknesses.

Other methods are no doubt possible. The above, however, comprise those which readily came to mind.

Comparison of Methods

A careful inspection of the above systems will reveal that there is no one method which at all times and in all places is superior. Each performs somewhat different functions and has somewhat different emphases. To ask which of the methods is better, accordingly, is analogous to asking the question, "Which is the better container for apples: boxes, bags, or baskets?" The only reasonable answer is "It depends on times, places, and circumstances."

In the practical situation the matter of motivation of the potential employee is a portion of the total value to be placed upon a method, as well as the excellence of the sieving which it produces. It may or may not be wise, for example, for a candidate to get the notion that he is "the one person in all the world able to fill this job," even though, in a certain very real sense, this may be true. Human experience is more in accord with the sentiment that there are many people who could do a given job; and conversely, that for every person there are many jobs in which he might do at least minimally well. In view of the fact that we know so little about the functioning of traits, let alone their organization, causation, and growth, it seems safe to say that the summation of the weighted scores method, perhaps in the general case, is closer to the realities of "human experience" than most of the

²² Or a similar formula in terms of β 's.

other methods cited, with the exception that when one is looking for unusual talent (with the expectation of hiring only one or two individuals at most from "among all the world") then the predominant or outstanding merit method is obviously superior. When employing the latter method if one is going to hire a very large proportion of all of the available candidates, then clearly one wants to be sure to refrain from getting any large aggregate of "weaknesses" and the minimum trait method is good. In this case it is understood that it is the production of the group rather than of a particular employee which determines the personnel bank-balance of the selector. If one is going to hire quite a few candidates but still a very limited number compared with the total number of applicants, the successive hurdles method will be economical and withal effective, providing the method is carried out in the approved fashion outlined above; namely, the administration of the most valid trait first, then of the next most valid, and so on down to the least valid last. When using this method Ruml's rank tangential coefficient²³ may be of aid in establishing the "critical scores." (Critical scores are in none too good repute among psychologists today. The concept grew up in a day when it was believed that failures were due to one "type" of humanity and successes to another, and that accordingly, if only one built the right kind of a test, he would secure a bimodal distribution. If only this were not untrue, it would be a very definite concept in view of the fact that it would be easy, at least comparatively, to determine what point should be set in order that the overlapping of distribution *A* on *B*, should be a minimum.) Where the numbers of cards to be sorted is "small" and the categories of traits are few in number, Keysort or Speedsort has several advantages, namely, that (1) a sorting needle is the only paraphernalia required and (2) some capital presumably may be made of the visual channels on the edges of the cards as successive sorts—preferably in a descending order of the weights or importances of the traits—are made. In this latter case an

²³ Ruml, Beardsley. "The Reliability of Mental Tests in the Division of an Academic Group." *Psychological Monographs*, Vol. 24, No. 4, Whole No. 105, 1917, p. 59 ff.

unbroken card across channels, the sorts being nearly completed, would indicate the case of a person below the minimum (how much can quickly be inspected on the card in question) on the category in question under observation.

There probably are other methods of selection. If so, it is hoped that this paper may encourage others to discuss the remaining methods and to compare their advantages and shortcomings with the methods herein outlined. Selection throughout the world—particularly in areas where so much leadership personnel has been destroyed by war—will be more important in the days to come than in any recent historical period.

•

THE COUNSELING PROGRAM OF THE VETERANS ADMINISTRATION

CARLOS E. WARD AND GWENDOLEN SCHNEIDLER
Vocational Rehabilitation and Education Service,
Veterans Administration

THE Veterans Administration is preparing to make the services of qualified counselors available to veterans throughout the nation. Counseling is being offered to all veterans who are eligible for vocational rehabilitation under Public Law 16, 78th Congress, or for education or training under Title II of Public Law 346, 78th Congress, and it is evident that a very large percentage of the veterans of World War II will be eligible for the benefits of these two laws which are being administered by the Veterans Administration.

Public Law 16, 78th Congress, approved March 28, 1943, may be regarded as a Disabled Veterans Vocational Rehabilitation Act, since its principal purpose is to provide vocational rehabilitation to overcome the handicap of disabilities which were incurred as a result of service in the armed forces during the period from September 16, 1940, to the end of the present war. Vocational rehabilitation will usually be attained through training provided for each veteran to fit him for employment consistent with the degree of disablement and suitable to restore his employability.

Public Law 346, 78th Congress, approved June 22, 1944, the correct title of which is "The Servicemen's Readjustment Act of 1944," became so widely known as the "G. I. Bill of Rights" prior to its enactment by Congress that many people still speak of it as the "G. I. Bill." Title II of this Act provides that veterans who meet certain eligibility requirements and whose education was impeded, delayed, interrupted, or interfered with by reason of entrance into service, or who desire a

refresher or retraining course, may be given education or training courses at approved institutions of their choice within certain time limitations prescribed by the Act. The veteran's length of service is an important factor in determining the duration of the education or training to which he may be entitled. Comparatively few veterans of World War II would fail to qualify for at least one year's education or training under Title II, if they should choose to apply.

When the fact is considered that complete counseling service, including modern techniques of testing and the use of compendia of systematized occupational information, is absolutely essential to the selection of an employment objective and of a training course suitable to the disabled veteran in need of vocational rehabilitation, it is evident that this alone would require an extensive Veterans Administration counseling program. But when to this is added the Veterans Administration's responsibility for counseling all eligible veterans who request educational and vocational guidance in connection with their applications for education or training under Title II of The Servicemen's Readjustment Act, then the magnitude of the task which this agency faces takes on new proportions.

While "educational and vocational guidance" is the term used in Public Law 346 to designate this service, the Veterans Administration has developed its counseling program with most careful regard to the fact that veterans will have greater assurance of achieving their educational objectives or occupational goals when mental conflicts, emotional maladjustments, and other types of personal problems are alleviated prior to or parallel with counseling and training. The plans and procedures of the Veterans Administration, therefore, are aimed at providing such thorough and complete counseling to each veteran claimant that he may be assisted in making the adjustments necessary for a useful life as a citizen. For this counseling service it is most important that competent, professionally trained, and otherwise well-qualified persons be employed.

For many veterans, guidance in the selection of an occupation or of an educational objective may be all that will be required. Other veterans may need assistance in handling

personal problems, in resolving mental conflicts, in attaining emotional stability, or in learning how to secure and hold employment. The Veterans Administration plans to furnish all types of counseling service which may be required in the case of an individual veteran. The veteran will also be guided in making intelligent use of other clinical and professional services, available to him through the Veterans Administration and other agencies, for the purpose of assisting him in making and maintaining the mental, emotional, and social adjustment essential to the attainment of his objectives. Each veteran is counseled in accordance with his needs as a person and educational and vocational guidance are not given without reference to the consideration of the other problems which affect the life of the individual.

The field organization of the Veterans Administration has been extended to carry out the general policies, plans, and procedures developed in the Central Office for counseling veterans. The program was started in 1943 by placing Vocational Advisers in the 53 regional offices in the different states. As the number of veteran claimants increase, Veterans Administration Guidance Centers are being established in a number of cooperating colleges, universities, and other educational institutions which have personnel qualified to render counseling services. This plan makes the counseling service accessible to veterans at points nearer their homes; it provides for the participation in this counseling service of a number of well-qualified people in colleges and universities who would not wish to sever their relationship with their educational institutions but who are able to contribute highly valuable service in the counseling of veterans; it affords veterans thorough and complete counseling in a suitable setting; and it enables the Veterans Administration to make arrangements, in advance, with educational institutions which, because they are going concerns, will be able to start counseling service to veterans on comparatively short notice when an increase in the number of claimants makes it desirable to have a larger number of Guidance Centers in any regional territory.

The fact that a veteran is requested to report for counseling

at a Veterans Administration Guidance Center located at any particular educational institution places no obligation upon him to take educational or training courses at that institution. Each Veterans Administration Guidance Center provides counseling service to any eligible veteran who may apply if he resides within the area to be served by the Guidance Center, and each veteran may take his training in any approved educational institution or training establishment where he can secure courses appropriate to the attainment of the educational or occupational objectives he has selected.

One of the most important advantages of the plan for establishing Guidance Centers in educational institutions is the provision it makes for a continuing supply of trained counselors, so that when the time comes to expand the service still further and to extend it from the nuclear points at the colleges and universities to the communities more remote from these it may be possible to make this expansion without reducing in any way the quality of the counseling service. Considering the probability that before long it may be necessary to provide counseling service at some 400 Guidance Centers at educational institutions and then to prepare to expand the service still more by supplying trained counseling personnel for additional offices of the Veterans Administration, it becomes apparent how important it is to make provision for a practical program of counselor training

For this purpose the Guidance Center plan which locates the practical work of counseling in the educational institutions which provide counselor training is in many ways ideal. Such colleges and universities not only are able to train counselors through their usual classroom instruction, but they are also in a position to combine with this instruction observation of actual counseling procedure. As the new counselors develop they can be given closely supervised practice in counseling techniques. Thus the established Guidance Centers not only provide veterans with the services of well-qualified and experienced counselors, but they also serve to increase the number of trained counseling personnel.

At each Guidance Center the Veterans Administration will

have at least one Vocational Adviser and one Training Officer. The members of the faculty of the educational institution who are assigned to counsel veterans at Guidance Centers are called "Vocational Appraisers" in order to distinguish them, in the records, from the Veterans Administration Vocational Advisers.

The Veterans Administration Vocational Advisers are selected from lists of professional counseling personnel prepared by the U. S. Civil Service Commission, and the counselors who are employed by the educational institutions to render service to veterans in Veterans Administration Guidance Centers have similar professional qualifications. The success of the counseling program depends upon securing counselors with the requisite professional training, who are technically qualified in psychology and in tests and measurements, who are experienced in performing counseling functions and who can work effectively in an organization following prescribed procedures.

The counseling principles, methods, procedures, and techniques used in the Veterans Administration counseling program are contained in the "Manual of Advisement and Guidance" which is being printed by the Government Printing Office, and these will not be described here. Short, intensive training conferences covering the specific procedures and techniques which are described in the manual will be conducted at selected locations from time to time as part of the in-service training program for those entering the Veterans Administration counseling program in order to supplement the basic background of training and experience which the adviser of veterans should have acquired over a period of years.

The counseling program of the Veterans Administration both in the regional offices and in the Veterans Administration Guidance Centers adheres to the policy that counseling does not meet required standards if it merely informs veterans of occupational and training opportunities without making a careful survey and a thorough analysis of the individual's education, work history, abilities, aptitudes, interests, and personality traits. In such an individual survey and analysis, extensive use of objective tests is required in most cases. It has been the policy of the Veterans Administration to administer only

those tests which the counselor prescribes for each veteran, rather than to make use of a standard battery for all. Counselors in Regional Offices and in Veterans Administration Guidance Centers select from well-standardized tests those which they desire to use and their selections are submitted to Central Office for review and approval. This plan has enabled the Veterans Administration, while initiating the counseling program, to utilize the best skills of counselors whose experiences in the administration of tests have been quite different, without waiting to train them all in administering a specified list of tests. As the program goes forward and as experience in the testing of veterans in connection with counseling is studied and evaluated, it is probable that the tests which serve a specific purpose in a large number of cases will be more and more widely used and the selection will thus tend to become more standardized.

To take a long-time view, the Veterans Administration's policy of providing the services of professionally trained counselors for veterans should increase the number of persons who are trained for counseling service so that, after the present emergency is past, such counselors may be available to carry on more extensive student personnel programs in universities, colleges, and secondary schools. It should also enable community agencies to render more counseling services to adults. A greater understanding both of the values and of the limitations of tests and of other counseling techniques may be attained through experience in this program. Another result which may be expected will be an increased demand for counseling services and for professional training for counselors. This training will probably take a more practical form by including greater opportunity for clinical in-service training as well as for theoretical work in psychology and related fields.

PERSONALITY TRAITS ASSOCIATED WITH ABILITIES. I. WITH INTELLIGENCE AND DRAWING ABILITY

RAYMOND B. CATTELL
University of Illinois

I. *Conception of the Research Problem*

It is customary to think of abilities as powers having an existence independent of other personality traits, i.e., of dynamic and temperament traits. In mathematical terms this is founded in the conception of unitary traits as factors (6). Since even different ability factors are mathematically independent of one another, it is not surprising that factors of different modality, e.g., temperament and ability traits, are still more confidently expected to be independent. Correspondingly, in clinical and general psychological terms the usual approach conceives of functionally independent traits or powers. Abilities, for example, are the tools of dynamic traits and may be used interchangeably by the same or different drives. Prediction of, say, the outcome of a son's antipathy to his father or a girl's overcompensatory concentration on school subjects rests first on an estimate of the strength of the drive but also on knowledge of the endowment in the various abilities which it may use.

The purpose of this paper is to show that the above analysis is only a first approximation to the truth. It proposes a more refined conceptualization and presents some new data, which, together with the data of an ensuing article (5), constitute a slight initial foundation for a factual edifice in this realm.

II. *The General Nature of Ability-Personality Trait Connections*

Clinically the connection of abilities with dynamic traits is often quite striking. Inferiority overcompensations sometimes

produce astonishing performances—relative to the individual's I.Q.—either in all school subjects or in some obscure field in which the individual finds he can at once jump ahead of, and avoid outright competition with, his rivals (1). The writer has also known two or three “idiots savants” whose outstanding “trick” performance in computation or memorizing could be traced to some initial accidental display, of comparatively slight eminence, by which the individual discovered that he could step into the limelight from the drabness of institutional life.

Psychoanalysis provides abundant instances of special skills, perceptual and motor, developed, like symptoms, out of the unconscious drives, relentlessly seeking expression. The complementary phenomenon of *conscious* drives, as sentiments, shaping abilities is so commonly realized as scarcely to demand illustration. Many of the special abilities distinct from intelligence isolated by research, e.g., Thurstone's Primary Abilities, may prove to be environmentally, dynamically shaped patterns, from general ability being impressed by particular investments of time and energy in certain conventional patterns of skills. (This aspect has been discussed more fully elsewhere (4) in connection with *the theory of fluid and crystallized abilities*.) The influence of major, overt sentiments upon ability patterns probably reaches its widest expression in respect to the self-regarding sentiment and it may be, for example, that the lesser mechanical aptitude of girls, and even the failure to find a mechanical aptitude factor among measurements on girls, will in the end be traced, not to difference of natural capacities as such, but to differences of dynamic adaptation stereotypes of the self-regarding sentiment.

Naturally, causal connections can run in either direction, or in a “causal circle”; and from general psychological observation a clinician would confidently say that examples of these three theoretical possibilities exist with almost equal frequency. Interests produce discriminatory and motor abilities as discussed above. But the individual who finds himself endowed with certain good natural abilities is likely to enjoy exercising them, and, in a competitive world, to find the dynamic pattern of his self-regard increasingly shaped by these abilities. Consequently, the establishing of connections between abilities and

personality traits is only the first—though a highly necessary—stage in investigation, needing to be followed by exploration for causal connections.

III. *Conditional and Wholistic Factors*

To dissect the present research problem into its ultimate roots one would need to ask whether the very distinction between ability, temperament, and dynamic traits may not be a chimera. Fortunately we have no need to enter upon so onerous an undertaking here, having carried it out elsewhere (7) and emerged with the conclusion that these common sense distinctions are real enough and capable of operational definition.

The psychologist familiar with correlation procedures is more likely to ask the very matter-of-fact question: "If abilities and non-cognitive personality traits sometimes correlate appreciably, how is it that the factors which have so far been discovered have been either pure ability factors, pure temperament factors, or pure dynamic factors?" The answer is, first, that there has been a tacit or unconscious conspiracy to maintain certain influences constant when giving tests, without mentioning—often without realizing—that such artificial conditions have been set up. We correlate ability tests given under conditions of quiet, of concentration, of common intention to do one's best. We correlate emotional responses in, say, nursery school children, observed under conditions in which cognitive abilities are not required in order to manifest emotion. That is to say, we always hold constant, in the measurement situation, the personality manifestations in which we are not interested.

Factors obtained under such test conditions we have named *conditional factors*. Since everyday life provides such constant conditions in a fair proportion of situations, conditional factors are by no means useless artificialities, when one comes to tasks of practical prediction.

Factors tend to be "pure" for a second reason, namely, that most investigators confine themselves to a deliberately narrow *sample* of the possible range of personality performances. They are looking for abilities, or even musical abilities only, or even ability to judge pitch alone. A very narrow range means less

communality, more unrelated specifics, fewer factors, and larger emphasis on the factor peculiar to that region.

The hypothesis can be put forward, therefore, that if one measured a collection of performances each under the naturally existing, varied conditions usually associated with the performances, and if one took an extremely wide sample of aspects of personality, many of the factors emerging therefrom would range at once over abilities, temperament traits, and dynamic traits. Factors obtained under such conditions we have called *wholistic factors*. Conditional factors will then appear as truncated forms of corresponding wholistic factors, cut down to their manifestations in a single modality.

General psychological considerations support such expectations. One can readily think of both *constitutional* and *environmental mold* source traits which should appear as wholistic factors. A single gene might be responsible for modifications in, say, certain parts of the midbrain, the cortex, the pituitary and the inner ear, causing simultaneous endowment in certain dynamic needs, some motor capacities, specific temperament traits, and particular auditory acuities. Similarly, particular school systems might frequently produce environment mold factors of covariation in certain acquired abilities, in forms of inhibition, and in specific interests and habits.

The exploration described below promised for the first time to reveal such wholistic factors if they exist; for it took a deliberately very wide array of personality aspects; in fact it sampled from what has been described in more detail elsewhere (2) as *the personality sphere*. Specifically it dealt with 35 clusters of traits, representative of all traits in the dictionary, rated for 208 adult males who had also been tested for intelligence, mechanical aptitude, drawing ability, verbal ability, mathematical ability, etc. The details have been described (3).

IV. *Personality Associates of Intelligence*

The principal personality correlate of intelligence known to psychologists is moral character and it is with respect to this correlation that the vast majority of past data has been gathered. The correlation is established through studies of intelligence of delinquents, alcoholics, etc., contrasted with nor-

mal controls, and through correlations of intelligence with rated or measured moral character qualities in a "normal" population.

This whole field has been very systematically and competently surveyed and summarized by Chassell (11). The normal-delinquent and other dichotomous data on the one hand and the continuous data on the other, when expressed in corresponding correlation form, agree very well, as also do the data from ratings and those from such actual conduct studies as those of Hartshorne, May, and Maller (14). Chassell (11) summarizes the results from many researches and several thousands of subjects as follows: "Expressed in correlational terms, the obtained relation may therefore usually be expected to fall between .10 and .39, and the true relation to be under .50." Since most of these researches are on groups of restricted range the figure for the population as a whole would be higher. Chassell finally estimates that the correlation between intelligence and the moral character qualities of self-control, unselfishness, reliability, industry, loyalty, etc., when corrected for narrow sample and attenuation, is close to .60. There are clear indications that the correlation is somewhat higher for children than for adults and, curiously enough, higher for the correlation with intelligence than for correlation with school achievement. This last, together with the fact that measured intelligence apparently correlates as well as rated intelligence, reduces the weight of the possible criticism that in some rating (teacher-child) situations the "goodness of character" might be a spurious product of pleasing authority by doing good school work.

The findings of the present research may be viewed from two angles (1) as straight correlations viewed in terms of correlation clusters or surface traits and (2) in terms of factors or source traits.

1. Rated intelligence, in this group of adults, was found to take its place in two clusters, as follows:

<i>Cluster Indexed as B3 in (2)</i>		<i>Cluster Indexed as B4 in (2)</i>	
(Items in descending order of correlation with remaining items)		(Items in descending order of correlation with remaining items)	
Original v .	Banal	Intelligent v . .	
Constructive . . v .		Clear thinking v .	Incoherent
Interests wide . . v . .	Interests	Logical ability . . . v . .	
	narrow	Given to reasoning . v .	
Independent . . . v .	Emotionally	Clever v . .	
	dependent	Spatial-visual ability . v . .	
Persevering . . v .	Quitting	Mathematical ability v	

In these clusters every variable correlates with every other above .40 and the mean r is about .75. Apart from various "cluster appendages" to these nuclei, provided by traits which correlate with some but not all of the items, these are the only two clusters into which intelligence enters. *B4* is essentially intelligence manifested as a cognitive ability surface trait. *B3*, which has also been found in the Sanford-Murray research, shows, on the other hand, intelligence amidst its closest personality associates.

2. In the factor analysis of the 35 clusters taken to represent all aspects of personality, one of the twelve factors—the second largest¹ in its contribution to the variance—appears to be a factor of general intelligence. Its loadings for all variables may be read in the published table (3). Here we shall discuss only traits with outstanding loadings. The six traits (each a cluster) with highest loadings are as follows:

Loadings in Personality Factor B²

(2)	.52	(Intelligent)	v	..	(Stupid)
		(Clear thinking)	.	v	..	(Incoherent, confused)
		(Clever)		v.	..	
(11)	.47	(Persevering)	.	v	.	(Quitting)
		(Painstaking)	v	.	(Slipshod)
		(Conscientious)	. . .	v.		(Conscienceless)
(4)	.43	(Thoughtful)	v	.	(Unreflective)
		(Deliberate)	. .	v	.	(Impulsive)
		(Austere)	v		(Profligate)
(28)	.42	(Stable emotionally)	. .	v	..	(Changeable)
		(Self-respecting)	. .	v		
		(Self-controlled)	. .	v	..	(Unselfcontrolled)
(12)	.41	(Intellectual)	..	v		
		(Analytical)	. . .	v	.	(Unreflective)
		(Wide interests)		v	.	(Narrow interests)
(3)	.41	(Independent)	v		(Emotionally dependent)
		(Reliable)	.	v		(Undependable)
		(Mature)	v	.	(Emotionally immature, irresponsible)

In the β rotation variable 28 falls slightly and 29 takes its place in the first six. 29 is

Alert	v	Absent-minded
Energetic	v		Languid
Quick		v	..			Slow

¹ It remains in this position, and substantially unchanged in character, in the alternative α and β factorizations which have been offered (10).

² This factor, which is well known as Spearman's " g " in ability studies, has been called " B " in the personality realm, to conform with the series of general personality factors and to express by alphabetical order its relative order of magnitude in that series.

The numbers at the left identify the cluster variables in the original list recorded in the factor analysis (3). When an actual intelligence test was correlated with the personality variables it confirmed the factor as being the ability factor of general mental capacity, outcropping in the personality realm, for it gave the following highest correlations, which agree substantially (5 of the 6 being the same) with those found in the factor analysis.

	r
(2) Intelligent, analytical, etc40
(12) Intellectual interests, etc.31
(11) Strong-willed, conscientious30
(29) Psychophysically vigorous, alert29
(3) Wise, mature, polished26
(35) Smart, assertive24

One may next attempt to glean the evidence resident in past researches as to the correlation between intelligence and individual traits in the realm of character—as distinct from total character—to see if the above order is confirmed. Below we have taken the researches of Terman, Webb, and others reported by Chassell and averaged the r 's by a method roughly correcting the result for different samples and for attenuated coefficients. However, the method does not justify recording actual correlation magnitudes and we merely give the rank order of the traits, the r 's of which range from about .7 to .3.

Cooperativeness (Highest)	(Average of 7 r 's)
Reliability, trustworthiness, responsibility	(" " 17 ")
Industriousness (mainly in school)	(" " 12 ")
Conscientiousness	(" " 10 ")
Sympathy	(" " 2 ")
Moral habits and ideals	(" " 6 ")
Unselfishness	(" " 2 ")
Sincerity (Lowest)	(" " 2 ")

Until some remaining evidence is presented we shall defer discussion as to whether this order agrees with that of the loadings in our B factor. The remaining evidence of the nature of the B factor is to be obtained from looking into the traits which the factor *does not* load, i.e., those in the hyperplane of the vector. This is a necessary enquiry in the recognition of any factor. The B factor had presented some difficulties in rotating for simple structure because it produced only a relatively faint and uncertain hyperplane: there were relatively few variables

with *C* (.32) and, less certainly, with *D*, *G*, and *I*. When second-order factors (9, 17) are calculated from the correlations between the primary personality factors one, and possibly two, second-order factors are found, loading factors *B* and *C* positively and more highly than any others.

Adequate discussion of the meaning to be given to the novel concept of second-order factors is not possible here. (See 9 and 17) The writer believes that the most likely explanation of the second-order factor loading *B* and *C* (Emotional Stability, inverse of General Emotionality) is that it is a social status factor, expressing the genetic adhesion (8) produced by assortative mating between intelligence, emotional stability, and other "success"-generating qualities that are intrinsically distinct.

So much for the sheer description of the connections between intelligence and personality structure. A fuller discussion of the possible interpretations and origins of these connections will be taken up towards the end of this article.

V. *Personality Associates of Drawing Ability*

About the personality associates of creative artistic ability, notably about the alleged "artistic temperament," much has been written outside scientific psychology. One could wish that our data might throw light on this matter, but actually it is restricted to drawing ability and may or may not have reference to total artistic feeling and creativity. One hundred and twenty-eight subjects, selected from the 208 adult males in the total personality research for more uniformity of age and educational background, were asked to draw a man sitting in a chair reading a book. The drawings, all done within a 20-minute period, showed marked variations in resource, originality, and wit but were rated, by art teachers, for *artistic drawing ability*.^a The reliability of the artistic ratings, as between the two judges, was .88.

Correlations with personality variables were carried out separately within each of the 8 groups of 16 within which the men were rated. The correlations agreed in sign in 6, 7, or all

^a "Correctness of drawing, proportion, quality of line, expressiveness."

of the 8 groups for the following variables: 9 (-), 11 (+), 13 (+), 21 (+), 22 (+), 23 (+), 25 (+), 28 (-), 29 (+), 30 (+), 31 (+), 32 (+), 33 (+), 34 (+) (see variable list in 3). On a population of this size any correlation above about .22 is significant at the 1% level. The following mean correlations (for 8 groups of 16) are therefore worthy of especial note from among those listed above.

Drawing Ability with:

(34)	Labile	v	Habit-bound	} $r = 29$
	Intuitive	v . . .	Logical	
	Careless of material things	v . .	Thrifty	
(31)	Sociable	v . .	Shy	} $r = 29$
	Responsive	v . .	Aloof	
	Hearty	v . .	Quiet	
(30)	Incontinent	v . .	Inhibited	} $r = 27$
	Gluttonous	v . .	Queasy	
	Curious	v . .	Unenquiring	
(9)	Grateful	v . .	Thankless	} $r = 26$
	Softhearted	v . .	Hardhearted	
	Easygoing	v . .	Short-tempered	
(33)	Tough	v . .	Sensitive	} $r = 25$
	Lethargic	v . .	Hurried	
	Talkative	v . .	Introspective	
(29)	Alert	v . .	Absent-minded	} $r = 24$
	Energetic-spirited	v . .	Languid	
	Quick	v . .	Slow	
(21)	Energetic-spirited	v . .	Languid	} $r = 24$
	Self-confident	v . .	Self-distrusting	
	Debonnaire	v . .	Aloof	
(22)	Responsive	v . .	Cold-hearted	} $r = 22$
	Genial	v . .	Brooding	

Because of the consistency of the correlation in all eight groups, despite its lowness, we suspect systematic connections also with (23) Exhibitionist, eloquent, flattering, (13) Easily jealous, self-pitying and (32) Optimistic

The agreement of this syndrome with the popular stereotype of the earthy, gay, spirited, unstable, careless, Bohemian artist is very striking, yet it must be emphasized that at the time the ratings were made no one had any idea that the purpose was to correlate them with artistic ability. The correlations are low, but there is a reason additional to the simple statistical test of significance for believing that they are real. This is that the pattern of traits corresponds strikingly to a functional unity, or small group of functional unities, already found

in the factor analysis of personality (3). Seven of the above eleven variables occur among the small group of variables highly loaded with and used for defining the *F* factor of Surgency v. . Melancholic Desurgency. Viewed more closely (see table of loadings, p. 88 of 3), this cluster of drawing ability personality associates is seen to be systematically produced by the *F* factor (Surgency) in collaboration with *H* (Charitable Rhathymia . v. . Obstructive Schizothymia) and to a lesser but perceptible extent with *E* (Dominance) and *J* (Vigorous, "Obsessional" Character). That is to say, if one picked out the variables high in these four factors, neglecting the influence of all other personality factors, and if one gave predominance to *F*, a slighter role to *H*, and a dash of *E* and *J*, the correlations between the members of this cluster would be fully accounted for. At this stage of research, however, no systematic, exact partialling out of the correlations has been attempted. The interpretation of this finding is discussed more speculatively below.

VI. *Interpretation of Correlations with Intelligence*

The correlating of intelligence with a complete range of personality manifestations confirms the earlier impression of narrower experiments that it correlates almost exclusively with what may be called "character" qualities and that the correlations are of substantial magnitude. Moreover, among the character qualities themselves the traits persevering, conscientious, self-controlled, reliable, emotionally independent, industrious, etc., correlate somewhat more highly than unselfish, emotionally stable, sincere. It looks as if intelligence is *directly* more associated with character conceived in a narrow, self-conscious sense, and with respect to habits that are acquired later and through conscious ideals, rather than with basic emotional integration and goodness of character in the wider sense such as might result from the emotional adjustment derived from the upbringing of the first few years or from some relatively constitutional stability. The above character associates also demonstrably include breadth of interests, habits of reflective thought, and analytical habits of approach to problems.

The association with the above restricted realm of character qualities may be interpreted either as the functioning of a single factor B (β factorization), or due to three further factors, G , J , and K , which are themselves correlated with B as factors (α factorization). The pros and cons of the alternative rotations which give these two possible interpretations have been discussed elsewhere (10), and until further research is done a choice is impossible. However, no matter which of these analyses is accepted, the present writer considers that the general psychological knowledge in this field favors the hypothesis that the character correlations are due to better intelligence leading to better learning, which would be expected to show itself in conduct situations almost as much as in academic situations. That is, the character patterns are "environmental mold" traits, patterns of reward and punishment in the culture, which "take" better on a basis of good constitutional "g" endowment⁴ than on poorer soil. At least during childhood it is an intelligent adaptation to develop good character. The converse hypothesis, that good character qualities lead to better performance in intelligence tests, can be given only a negligible role, in view, for example, of the negligible effect which conative variations of all kinds have been shown to have on intelligence test performance.

Although character qualities of the above restricted kind are *directly* to be connected with intelligence, the character qualities (for such they are commonly considered) of the C factor, which we may call deeper emotional integration or stability (see full description in 3), are connected only indirectly, by a second-order factor, saturating both C and B , and causing their appreciable correlation. We are inclined to think that most second-order factors will turn out to be best explained in causal terms as a common cause influencing both the correlated variables (i.e., as the third of the logically possible alternatives in explaining correlation).

⁴ That these correlations run a little lower with adults may be due to the fact that years of trial and error experience compensate and reduce the learning gains due to intelligence alone. After all, much conduct learning is blind trial and error, even for the more intelligent. A radically different alternative would be that intelligent adults more quickly unlearn the moral habits taught them as children, but this would assume that intelligent people more frequently consider moral habits undesirable.

The *C* and *B* connection arises, we believe, from the effect of family and of social status (a) environmentally, in that more intelligent children, having more intelligent parents, will tend to get wiser handling of those emotional problems of the early years of childhood which determine the deeper emotional adjustment (c) and (b) genetically, by reason of what has been called the social status orientation⁵ of genetics patterns (9) showing itself in *socially conditioned genetic adhesions*. In this case, we presume, both high intelligence (positive *B* factor) and high emotional stability (positive *C* factor) are selected for social promotion and become genetically linked by assortative mating within social classes.

VII. *Interpretation of Drawing Ability Correlations*

Previous studies of drawing ability⁵ as an artistic performance, as distinct from Goodenough's ingenious use of it to measure intelligence (14), have on the whole failed to give us any clear picture of its ramifications as an ability or its connections with artistic abilities in general. Meier's survey (16) of the problem stresses that emotional and temperamental qualifications are as important as those connected with mere skills. Tiebout's (18) research, unfortunately on few cases, found no apparent superiority in various motor skills among those superior in drawing, but found superior power of observation and ability to retain visual impressions, for days or for months. Drep's (12) findings point the same way, showing better memory for visual form but not better motor ability, and indicating greater emotional sensitivity and more neurotic tendencies among abler artists than among others.

Artists familiar with the life of outstanding artists of the past and present were inclined to consider Van Gogh, Gauguin, Toulouse-Lautrec, or Whistler more typical than the staid Giotto and to impute higher general emotionality, impulsiveness, and ability to see everyday things freshly (through greater lability), and more imagination generally, to the true artist. One might add the consideration that psychotics have some-

⁵ Many studies have unfortunately been on artistic ability as a whole, begging the question as to whether several distinct talents may not be involved.

times been known to paint for the first time after the onset of the psychosis, and that several instances are known, e.g., Churchill and Hitler, of vigorous minds turning to painting in times of political or professional failure. In short, the present writer would argue that any researcher in the future who makes statistical analysis of life *in situ*, i.e., of actual artists is likely to find the personality traits associated with artistic performance, as defined by these more intensive researches, amply illustrated.

Common observation also seems to indicate that there is a fairly powerful hereditary influence in drawing ability. On the bases of the factor associations found here we would argue contingently that it is largely a factor for the temperamental tendencies which is inherited and that the drawing ability *per se* is acquired on the basis of the interests which these generate.

The *F* factor of surgent temperament (associated in its extreme loadings with conversion hysteria), with its test manifestations of "fluency" and "imagination" which have long since been demonstrated to be part of it (1), seems to be the first condition for the development of drawing ability. The cyclothyme-like factor *H*—which, incidentally, is one of the two factors loading "aesthetic interests" in the original factor analysis—is the second factor having correlation with drawing ability and perhaps corresponds to the element of more passive sensitivity and appreciation in the interests which lead to drawing skill. The connection of cyclothyme temperament with appreciation of color and the visual arts was stressed by Kretschmer (15), who considered the high hereditary incidence of cyclothyme constitution in the Alpine-Mediterranean racial regions to account for their pre-eminence in production and appreciation of visual art. The indicated correlations of *E* (Dominance) and *J* (Vigorous, Obsessional Character) are too slight to justify discussion until further explored.

If clinical type observation, aided by the definiteness of factors engendered by factor analysis, may be admitted as a guide to further research, the present writer would suggest that the following associations are likely to become important when total, creative artistic ability, as distinct from drawing ability alone, is studied. First, observation suggests that the loadings

of the cyclothyme factor (*A* or *H*) will be found almost as important as those of Surgency (*F*). Secondly we can surely expect that factor *C*, in its negative loadings, will have some role. This factor connotes the instability, plasticity of thought and purpose, richness of emotionality and freshness of emotional approach, which the world evaluates one way and artists another. There is no hint of this factor in our results because we did not deal with creative artists but with soldiers selected for emotional stability and measured with respect to straight drawing ability. However, we suggest that it would be desirable to confirm the above personality factor associates of drawing ability before directing research to the more complex personality problem of artistic creativeness.

VIII. Summary

Intelligence. Intelligence appears as a general factor (*B*) among personality traits, loading particularly *character* traits, and notably those good habits which may be consciously acquired. This factor correlates, however, to the extent of about .3, with a distinct factor (*C*) of emotional stability and integration, and together with other factors they yield a second-order factor, which may be the genetic adhesion of intelligence and temperamental (emotional) stability produced by social stratification. Past findings fit in well with these factorial interpretations.

Drawing Ability has significant correlations with eight out of the 35 surface traits used to represent the total personality sphere. These can be best explained as due to low positive correlations of drawing ability with Surgency (*F* factor) and Rhythmic Cyclothymia (*H* factor), and possibly slighter correlations with Dominance (*E* factor) and Vigorous Character (*J* factor). This personality pattern very distinctly resembles that observed in well-known artists, but it is suggested that total artistic ability, as distinct from artistic drawing ability alone, is also likely to involve General Emotionality (Negative *C* factor).

The writer wishes to express his thanks to Thelma Alper and Virginia Carvell for expert help in completing the drawing section of this research.

REFERENCES

1. Cattell, R. B. *A Guide to Mental Testing* London University of London Press, 1936
2. Cattell, R. B. "The Principal Trait Clusters for Describing Personality" *Psychological Bulletin*, XLII (1945), 129-161
3. Cattell, R. B. "The Description of Personality. III. Principles and Findings in a Factor Analysis." *American Journal of Psychology*, LVIII (1945), 69-90
4. Cattell, R. B. "The Measurement of Adult Intelligence." *Psychological Bulletin*, XL (1943), 153-193.
5. Cattell, R. B. "Personality Traits Associated with Abilities. II. With Verbal and Mathematical Ability." *Journal of Educational Psychology* (In press.)
6. Cattell, R. B. "Personality Structure and Measurement I. The Operational Determination of Trait Unities." *British Journal of Psychology*. (In press.)
7. Cattell, R. B. "Personality Structure and Measurement. II The Determination and Utility of Trait Modality." *British Journal of Psychology*. (In press.)
8. Cattell, R. B. "The Cultural Functions of Social Stratification. I Regarding the Genetic Bases of Society." *Journal of Social Psychology*, XXI (1945), 3-23
9. Cattell, R. B. "Oblique, Second Order and Cooperative Factors in Personality Analysis." *British Journal of Educational Psychology*. (In press.)
10. Cattell, R. B. "Simple Structure in Relation to Alternative Factorizations of the Personality Sphere." *Journal of General Psychology*. (In press)
11. Chassell, C. F. *The Relation between Morality and Intellect*. New York. Teachers College Bureau Publications, 1935.
12. Dreps, H. F. "The Psychological Capacities and Abilities of College Art Students of High and Low Standing" *Psychological Monographs*, XLV (1933), 134-146.
13. Goodenough, F. L. and Anderson, J. E. *Experimental Child Study*. New York: Appleton Century, 1931.
14. Hartshorne, H, May, M A, and Maller, J. B. *Studies in Service and Self Control* Vol II of *Studies in the Nature of Character*. New York: Macmillan, 1930.
15. Kretschmer, E. *Physique and Character*. Berlin: Springer, 1929.
16. Meier, N. C. *Art in Human Affairs: an Introduction to the Psychology of Art*. New York: McGraw Hill, 1942.
17. Thurstone, L. L. "Second-Order Factors." *Psychometrika*, IX (1944), 71-100.
18. Tiebout, C. "The Psychological Functions Differentiating Artistically Superior from Artistically Inferior Children" *Psychological Monographs*, XLV (1933), 108-133.

FACTOR ANALYSIS OF OCCUPATIONAL APTITUDE TESTS

Staff, Division of Occupational Analysis, War Manpower Commission

OCCUPATIONAL research projects have been conducted by the Division of Occupational Analysis since 1934, when it was known as the Occupational Research Program of the United States Employment Service. One of these projects has been the construction of aptitude tests. Batteries of aptitude tests have been validated for various occupations and are used in local offices of the Employment Service to aid interviewers in selecting the most satisfactory beginners for referral to jobs or training courses. In developing an aptitude-test battery for an occupation, a number of aptitude tests are "tried out" on employed workers (or trainees) for whom objective criterion data can be obtained, and the best combination of tests constitutes the battery.

The Division has recently been conducting a series of factor analysis studies to determine the important factors in its aptitude tests. Interest in the problem of the basic factors or fundamental aptitudes underlying the Division's occupational tests has been heightened by the recent emphasis on the counseling approach in vocational placement and guidance programs. From a practical point of view, the preparation of separate selection batteries for 20,000 different occupations is an unsurmountable task. However, if occupational proficiency can be expressed in terms of a few relatively independent aptitudes which, in various combinations, account for the differences in job aptitude requirements, then the use of tests for general counseling purposes becomes feasible. Several factor analysis studies have therefore been conducted in order to isolate the basic aptitudes measured by the Division's tests and to select a small number of measures of these factors for com-

bination into a counseling battery. It is planned to standardize this battery for a large number of occupations which will then be classified into groups on the basis of similarity of aptitude requirements.

The Experiment

Several experimental batteries of tests were administered to a total of 2,156 persons for the factor analysis studies. The data are divided into nine experimental groups. The number of subjects, the number of tests, the geographical location, and the factors identified for each group are shown in Table 1. The smallest battery consisted of 15 tests; the largest, of 29. There was a great deal of overlapping of the tests among the several batteries; but, in all, some 59 tests were subjected to analysis. Some of the factors found are well known and are similar to those which have been discussed by Kelley, Thurstone, and other investigators. Others have not received attention heretofore.

In group 0, 19 tests were administered to 1,079 male applicants for defense training courses in Erie and Pittsburgh. The age of the subjects ranged from 17 to 39 years, with a mean of 23 years, and all had completed at least six years of education.

TABLE 1
Description of the Experimental Groups

Group	Number of Subjects	Number of Tests*	Location	Factors
0	1079	19	Erie and Pittsburgh	O S P Q A T F M
1	221	25	Dallas and St. Louis	O V N S P Q A T F M
2	99	29	Sacramento	O N S P Q A T F M
3	141	15	West Virginia	O N S P Q T F M L
4	138	25	Philadelphia	O N S P Q T L
5	275	27	Cincinnati, Detroit, Cleveland, Toledo and Chicago	O V N S P Q A T F M
6	98	28	Chicago	O V N S P Q A T F M
7	594	25	Composite of Groups 1, 5 and 6	O V N S P Q A T F M
8	204	24	Same as Group 2†	O N S P Q A T

* Some of the correlational matrices were actually larger than here indicated, because of the inclusion of age and education as additional variables.

† This group includes all of Group 2 plus additional subjects from the same training course for whom data on five tests were not available.

In the remaining eight experimental groups, all of the subjects were males and most of them were trainees enrolled in Vocational Education National Defense Training courses. The age of the subjects ranged from 17 to 39 years, with a mean of 28 years. The mean number of years of education completed was 11, and ninety-nine per cent of the subjects had completed between 8 and 16 years. It is estimated that about five per cent of the sample were Negro, and the rest white.

General Description of the Tests

A total of 59 different tests were employed in the several factor analysis groups. Fifty-four of these, 48 paper-and-pencil tests and six apparatus tests, were constructed by the Division. The other tests were the *O'Rourke Survey Test of Vocabulary (Form X4)*, the *Revised Minnesota Paper Form Board* (Likert and Quasha), the *Minnesota Spatial Relations Test*, the *Minnesota Manual Dexterity Test—Placing*, and the *Minnesota Manual Dexterity Test—Turning*. In addition to the tests, age and education were included as variables.

In general, the tests constructed by the Division are speed tests, with time limits for the most part in the neighborhood of five minutes. The individual tests are homogeneous in content, since it was planned to combine them into batteries. The fact that they are intended for use in offices of the United States Employment Service, chiefly for industrial workers, is a central consideration. There was no need to construct many verbal or "intellectual" tests. Emphasis has been placed instead on development of tests of perceptual and spatial ability and of dexterity. It was intended to construct tests that appeared to have validity for occupations, but, on the other hand, were not so analogous to specific jobs as to impair the applicability of the tests for widespread use. All the tests are so constructed that they can be easily administered by personnel without extensive technical training.

The Factors

Thurstone's methods (2, 3) were employed to extract the centroid factors from the correlational matrices and to rotate

them to a meaningful structure. For each group, a solution was first obtained which satisfied the criteria of simple structure. The effect of maximizing the number of zero loadings on as many factors as possible is equivalent to explaining test performance by a minimum number of factors. Simple structure is essentially the factor analysis analogue of the doctrine of parsimony.

It was discovered in each group that the first solutions had very nearly orthogonal structures. In one group, for instance, the largest correlation between factors was .099. The factors in an orthogonal structure are entirely independent and uncorrelated; when the factors are correlated among themselves the structure is said to be oblique. Since the structures were very nearly orthogonal, and inasmuch as the solutions were not so exact that different investigators would have obtained identical correlations between the factors, it was decided to impose an orthogonal structure on each group and the rotational process was continued until this was achieved. There is an important advantage to the final solutions so obtained: Comparisons of the results are rendered less ambiguous, in that reference can be made to factors which bear an identical relation to all other factors in each group.

The smallest number of common factors established in any group was seven, and the largest was ten. In all, eleven different common factors were found. Preliminary results from each group were applied to all the others, so that in a few groups more factors were determined than could be justified in any one of these if conducted independently. Consistent results were obtained from the several correlational matrices, in that the factors common to a related group of tests could always be demonstrated regardless of the composition of the remainder of the experimental battery. The loadings of a factor on a test for different groups varied to about the same extent as correlations for identical pairs of tests in the different groups.

Among the factors most readily established were the verbal (*V*), numerical (*N*), and spatial (*S*) factors. Although only three verbal tests appeared in the experimental batteries, the fact that their loadings ranged as high as .58 and did not fall

below .40 serves to identify this factor. One of these tests is the *O'Rourke Survey Test of Vocabulary*. Another is a proverbs test which measures the ability of the subject to understand proverbs. The third is a test of antonyms and synonyms.

The numerical factor (*N*) was found to be present in a wide variety of numerical tests: A test of arithmetic fundamentals (addition, subtraction, multiplication, and division), a specialized test of decimals and another of fractions; and reasoning problems, in which the subject solves verbal arithmetic problems. Two speed tests in one-digit arithmetic, one in addition, subtraction, and multiplication, and the other in addition only, are also classified as numerical tests.

The spatial factor (*S*) appears in about a dozen of the Division's paper-and-pencil tests, and was also found in the *Revised Minnesota Paper Form Board* and the *Minnesota Test of Spatial Relations*. The Division tests consist of two- and three-dimensional figures. Examples are surface development in which a pattern is matched with a three-dimensional object constructed from it, a picture test of the assembly of mechanical objects, another on fitting together abstract geometrical objects, and a test in selecting the mirror image of a line pattern. Thurstone remarks on the visual character of his spatial factor (4, 79-80) and the space tests in these studies may be similarly described. Kelley's description (1, 10), "manipulation of spatial relationships in so far as independent of differences in visual acuity," is also applicable. The existence of a perceptual factor tends further to limit the spatial factor. The large number of tests described above, all with substantial projections on this factor, make it easily identifiable as the same space factor found by other investigators.

One factor found in these studies presents difficulties in interpretation. This factor was found in each of the groups and is present in significant amount in about two dozen tests. The tests which have significant projections on this factor include all of the verbal tests, all of the numerical tests except the two speed tests of one-digit arithmetic, and almost all of the spatial tests. The factor was also present in a letter series test, a word memory test, and a perceptual relations test; this is in-

teresting because none of these tests have significant projections on either V , N , or S . It appears to have some of the properties of Spearman's G , but the two-factor theory has no place for group factors like V , N , or S . On the other hand, this factor has a wider significance and is more persistent than either Thurstone's R or I (4, 86-88; 5, 158-159). It appears to possess many of the properties that teachers, test examiners, and clinical psychologists would attribute to "intelligence." In Table 1 this factor has been designated, noncommittally, as Factor O .

A matter of interest, relating to current theories of general intelligence, is that this factor has been established in a sample of adults, ages 17 to 39. This tends to dispose of some theories that this factor could be established only among children, and that it amounts to a common maturational factor. The projection of age on this factor in this study is consistently negative, being about $-.250$. The loading of V on age, on the other hand, is consistently positive, about $.320$. From this one could possibly conclude that older individuals have greater facility in expressing themselves regarding familiar situations, and that younger individuals have greater facility in coping with new situations.

Among the remaining factors found are two perceptual factors, P and Q . P is present in a test on matching figures of various sizes and shapes, in a test on matching shaded figures, in a test on distinguishing figures which differ slightly in length, width, area, size of angle, or degree of curvature, and in many of the tests which also measure spatial ability. Altogether it appears to be present in over twenty tests. The factor Q is present in about nine tests, including a name comparison test and a number comparison test similar to those in the *Minnesota Test of Clerical Ability*, the arithmetical decimals test, a coding test, and a test on number copying. It is also found to some extent in several of the other tests which measure verbal or numerical ability. It may be further noted that the test of number comparison appears to measure P as well as Q and that the test of name comparison appears to measure O as well as Q . Another comparison test is one in which pairs of geo-

metrical figures are determined to be the same or different. Results are not conclusive but indicate that this test measures P and not Q .

The essential difference between the two perceptual factors is not entirely clear. For the most part the tests measuring P involve geometrical figures while the tests measuring Q involve words and numbers. Another way of interpreting the difference is to say that the tests which measure Q deal with materials emphasized in formal education, while the tests which measure P deal with materials to which little or no school time is devoted. The loading of P on age is about $-.400$; the loading of Q on age has not been positively established but is much nearer zero. Education has positive projections on both P and Q , but it is much higher on Q .

An aiming factor (A) was found in five paper-and-pencil adaptations of standard laboratory tests. In one test, the subject crosses the bars of a series of letter H's without touching the sides of the letter. In another test, the subject is required to make line strokes in squares. Apparently what is involved is accuracy or precision of movement. Whipple (6, 147-151) has applied the term "aiming" to various laboratory tests which measure accuracy of movement.

In the process of establishing orthogonal structures the speed factor (T) was found. It was discovered in one of the first groups that the factor A could not be established as a factor orthogonal to all the others. Leaving it in an oblique position would signify that the factor A , as then established, measured something also measured by the other factors. It was hypothesized that what was being measured in common was quickness or rate of movement, or speed. This was undoubtedly a general factor in a classical sense, since all the tests under study are speed tests. A general factor cannot be found directly by simple structure, and therefore the speed factor was established arbitrarily as a factor orthogonal to all the others. With the introduction of this factor, A was easily established in an orthogonal position without violating simple structure.

The factor T has less consistent loadings than any other factor, but approximately forty tests have significant projec-

tions on it in one or more groups. In general, it is easier to establish the existence of a phenomenon when it is sometimes present and sometimes absent than when it is always present. Related to this difficulty of establishing qualitatively the existence of this factor is undoubtedly the lack of consistent quantitative results. Among the tests with the highest loadings on *T*, however, are consistently those with moderate loadings on *A*. From this it follows that an increase in precision in the aiming tests results in a decrease in speed. A new speed test was constructed which demands practically no precision, and another is under construction demanding a large amount of precision. These tests will be included in future factor studies.

Two dexterity factors resulted from the apparatus tests. These have been designated *F* and *M* for finger dexterity and manual dexterity, respectively. An interesting finding was that the *Placing* and *Turning* parts of the *Minnesota Test of Manual Dexterity* had practically an identical factor composition, being almost pure tests of manual dexterity. Substantially the same finding obtains for Parts I and II of the *Peg Board Apparatus* developed by the Division. In Part I pegs are moved from one set of holes to another set of holes using both hands simultaneously, and in Part II the pegs are transferred from one hand to the other and replaced in the hole upside down. Another apparatus test involving fine assembly work measures *F*. However, disassembly of the same parts is a composite of *F* and *M*. *F* is significantly present in five tests and has moderate loadings on two others; and the existence of *M* is readily established from its presence in significant amount in five tests.

A factor *L* was tentatively established in two of the factorial studies, for four or five different tests. One of these is an analogies test consisting of line drawings and another is a verbal test in following directions. All of the tests with significant projections on this factor require the solution of problems by formal rational processes. This factor is a narrow reasoning factor and is being called the Logic factor. Age has a loading of about $-.350$ on *L*, and Education has a loading of $+.350$ on *L*.

It had been originally intended to include both a word

memory and a picture memory test in order to measure a possible memory factor. Unfortunately, only one memory test was administered, and no memory factor was found

Discussion

Group factors have great significance for vocational counseling. This point may be developed briefly. If only a single general factor accounts for vocational success, all that one could do is establish a hierarchy of jobs, in order of general ability, as was done with the Army Alpha results after the last war. It would not be possible to distinguish, for instance, a potential lawyer from a potential engineer, assuming that engineers and lawyers possessed on the average an equal amount of general ability. If we assume, on the other hand, that engineering requires more ability than law, then it would follow that *every* individual would find law easier than engineering.

On the other hand, if one were to postulate that every task required its own separate specific ability, it would be impossible for the vocational counselor to assess all these abilities, and to predict all the vocational aptitudes the applicant might possess.

With group factors perhaps only a few hours of testing are required to sample the significant aspects of behavior. Since it is unlikely that each occupation requires a different set of aptitudes from that of every other occupation, those occupations with similar requirements could be grouped together into fields. Then, on the basis of a relatively small number of test scores, prediction could be made for a number of occupations.

REFERENCES

1. Kelley, T. L. *Crossroads in the Mind of Man*. Stanford University: Stanford University Press, 1928.
2. Thurstone, L. L. *The Vectors of Mind*. Chicago: University of Chicago Press, 1935.
3. Thurstone, L. L. "A New Rotational Method in Factor Analysis." *Psychometrika*, III (1938), 199-218.
4. Thurstone, L. L. *Primary Mental Abilities*. Psychometric Monographs, 1938, No 1.
5. Thurstone, L. L. "Experimental Study of Simple Structure." *Psychometrika*, V (1940), 153-168.
6. Whipple, G. M. *Manual of Mental and Physical Tests*. (2nd Ed.) Baltimore: Warwick and York, 1914. Part I.

THE INTERESTS OF FOREST SERVICE MEN

EDWARD K. STRONG, JR.

Stanford University

THE top executive who has worked up from office boy has a profound belief that others could do likewise if they only would try hard enough. Is this true? Is willingness to work the major consideration or must abilities and interests also be taken into account? Suppose every nonsupervisory position was filled with a well-qualified man, would that insure a sufficient number for promotion to supervisory positions as vacancies occurred?

Many organizations, both public and private, maintain the policy of recruiting top management by promotions from below. Other organizations select men specifically for executive and administrative work and promote relatively few upwards from below. Personnel practices relative to selection and training reflect the policy of the organization involved. In far too many cases, however, no one in authority has ever definitely decided what the policy should be nor are the facts available upon which such a policy should be formulated.

Some data are given below which suggest that in the forest service the interests of district rangers and administrators differ so appreciably that it is questionable whether many of the former will or should be promoted to management. And it is furthermore questionable whether there are enough men in the lower bracket with the interests of administrators to supply the service with properly qualified men at the top.

In 1936, through the cooperation of the United States Forest Service, 410 members of that service filled out the *Vocational Interest Blank*. On the basis of these blanks an interest scale was developed to measure the degree to which a man has the interests of forest service men rather than the interests of men in general.

Such blanks may also be scored so as to reveal how similar

a man's interests are to the interests of men in 35 other occupations. Thus, it is possible to say that a particular ranger has more the interests of an engineer than of a forest service man and to say that a supervisor has not only the interests of a forest service man but also to a lesser degree the interests of a production manager, a personnel manager, a public administrator, and an office man. Such a supervisor might be expected to function somewhat differently from one who had an equally high forest service interest but whose secondary interests were those of an engineer, farmer, aviator, and policeman.

The data in this report are based upon the original group of 410 men plus about 50 additional cases of supervisors and regional and Washington administrators, obtained in 1941 in connection with a study of public administrators, financed by the Committee on Public Administration of the Social Science Research Council.

The data pertain to interests—what a man likes and also dislikes to do. The data do not include measures of general or specific abilities. Interests and abilities must both be considered before a final answer to the questions before us may be obtained. But what a man is interested in does play an important role, and it is this aspect of his behavior with which we are concerned here.

Scores of Forest Service Men on the Forest Service Interest Scale

As stated above, the forest service interest scale is based upon the records of 410 men, whose average age in 1936 was 38.5 years. Approximately half were district rangers; all but three of the remainder were assistants to supervisors, assistant and associate supervisors, and supervisors. In terms of rank the criterion group averages just above that of district ranger. Forest service interests as measured by the scale represent the interests of men from district ranger to supervisor.

Table 1 gives the forest service interest scores of 430 men, distributed by age and rank in the service.¹

¹ The classification of forest service men and the follow-up referred to below were made by Paul P. Pitchlynn, formerly assistant regional forester at San Francisco. The 410 blanks constituting the criterion group were also obtained largely through his efforts.

TABLE 1

Number of Forest Service Men by Age and Rank, and Mean Forest Service Interest Score of Each Sub-Group

		Age								Total
		25	30	35	40	45	50	55	60	
District	N	37	38	40	30	24	13	7	.	189
Rangers	M	53.2	52.5	51.5	47.5	46.0	47.9	38.2	.	500
Assistant to	N	11	3	3	1	.	18
Supervisor	M	47.8	60.7	45.3	39.0	...	49.1
Assistant	N	10	28	18	4	9	3	1	.	73
Supervisor	M	54.1	52.7	52.1	40.5	46.2	46.3	37.0	.	50.8
Associate	N	1	6	4	3	.	1	.	.	15
Supervisor	M	40.5	51.7	47.8	50.0	.	54.0	.	.	49.7
Supervisor	N	...	18	17	11	30	11	12	...	99
	M	.	47.5	47.5	51.1	48.3	48.9	41.3	.	47.5
P-6*	N	1	4	10	2	3	20
	M	50.0	50.0	44.2	31.0	37.3	43.3
P-7-P-8	N	1	1	4	8	2	16
	M	43.0	50.0	39.5	30.6	30.5	34.8
Total	N	59	93	82	50	68	42	31	5	430
	M	52.1	51.8	50.4	47.8	47.3	46.5	37.0	34.6	48.6†

* P-6, P-7, and P-8 refer to grades of positions in the professional service of the Federal government, many of which entail a considerable amount of administrative responsibility, and with respective base salaries of \$5600, \$6500, and \$8000 (as of the time of the administration of the test)

† Only 3 of the P-6 and none of the P-7 or P-8 personnel were included in the criterion group. If they were excluded and 13 cases, not so far classified, were included, the mean standard score of 410 criterion cases would be 50.

The data make clear that such scores decrease with age for all seven ranks in the service. The decrease in score amounts to only 4.8 from age 27.5 to 47.5 years, whereas it amounts to 15.7 from age 47.5 to 62.5 years.

Decrease with age in score on the forest service interest scale is a phenomenon strikingly peculiar to the profession. Data similar to those in Table 1 have been published for 29 occupations.² The average decrease in score for these 29 occupations, including forest service, amounts to 1.4 between the ages of 27 and 57 years, whereas the decrease is 15.7 for forest service men. There are appreciable decreases in score amounting to 10.5 for personnel managers, 7 for aviators, 5 for realtors and physicians and 4 for life insurance salesmen. The only occupation exhibiting increase of score on its own scale is that of minister, where

² E. K. Strong, Jr. *Vocational Interests of Men and Women*. Stanford University Press, 1943, Table 75

the increase amounts to 6.9 score points. For most occupations age affects interest scores very little, but for some reason the reverse is the situation in the forest service group.

This phenomenon explains in part why a few successful forest service men score low in the forest service interest scale. They score low because they are older men and all older men average considerably below 50.³

The question naturally arises as to why forest interest scores decline with age among forest service men. Three possible explanations may be considered: first, such interests actually decline with age, second, there has been a change in the type of men entering the service and the younger men, constituting the larger proportion of the criterion group, have largely determined the norms for the group; and third, men with high scores leave the service in later life to a larger degree than men with lower scores.

The preponderance of data in our possession does not support the first explanation, that the kind of interests possessed by forest service men naturally decline with age. For example, when samples of 25-, 35-, and 45-year-old men drawn at random from the population are scored on the forest service interest scale the mean scores are, respectively, 28, 27, and 31, indicating a slight rise in score with age.⁴

How about the second explanation? In the early years of the forest service few men were college graduates, whereas in recent years many have been recruited from colleges, particularly from schools of forestry. Has this change in selection affected the interests possessed by forest service men?

The relationship of age to amount of education among 187 district rangers is as follows:

³ The standard score of 50 is the average obtained by forest service men. For convenience scores are often expressed by the three letter ratings of A, B, and C. An A rating (scores of 45 to 75) means that the individual has the interests of men successfully engaged in the occupation; a C rating (scores below 30) means he does not have such interests, and a B rating (30 to 44) means he probably has those interests but we cannot be as sure of the fact as in the case of A ratings. Note that a C rating does not say a man is not interested in a particular occupation, it says he does not find interesting a whole range of activities which successful men in the occupation find interesting.

⁴ *Ibid.*, p. 272

<i>Age</i>	<i>Number</i>	<i>Years of education</i>
20-29 years . . .	38	16.1 (college graduation)
30-39 years . . .	77	13.6 (nearly 2 years of college)
40-49 years . . .	54	10.2 (2 years of high school)
50-59 years . . .	18	9.0 (1 year of high school)
Total	187	12.7 ((2/3 year of college)

On the basis of these data one might surmise that the decrease in forest service interest scores with age is attributable to lesser amounts of education. If, however, the age factor is partialled out we obtain the following interest score ratios for the six subgroups in terms of amount of education:

6 to 8 years, grammar school	83
1 to 2 years, high school	94
3 to 4 years, high school	107
1 to 2 years, college	95
3 to 4 years, college	101
1 to 2 years, graduate work	105

There is here no indication of increase in forest service interest score with increased amounts of education, except that those with only grammar school education do score lower than the remainder. But there are too few cases in this category to warrant emphasis upon the exception. Among supervisors there is no difference in forest service interest score between 24 college and 26 noncollege graduates. Apparently increase in amount of education has not affected forest service interest score. It is still possible that there have been other changes in selection besides education and that these undisclosed factors contribute to decrease in score with age but we have uncovered no proof of this.

The third explanation, that forest service scores decline with age because older men with high scores leave the service to a larger degree than those with low scores, can be established only by a long time follow-up of the men so far tested on the *Vocational Interest Blank*. Evidence presented below indicates that promotions, on the average, go to men with lower interest scores in mechanical pursuits and higher scores in general administrative interests. Such a hypothesis is tenable as far as the relationship of forest service and public administration interests is concerned, for they correlate only .21, which means that some men with high interest in one of these fields will have low interest in the other. It is therefore possible that some

men with high forest service interest scores have left the service when they failed to receive a promotion and that there have been enough such cases to explain the decrease in average scores with age.

Occupational Interests of Forest Service Men

So far we have considered only how far forest service men score on the forest service interest scale. Consider now how they score on other occupational interest scales.

Table 2 gives the percentage of rangers, supervisors, and administrators who rate A and B+ (scores of 40 to 75) in interest for 36 occupations. Such percentages indicate how many forest service men definitely have the interests of men in those occupations. For example, 88 per cent of rangers rate A and B+ on forest service interest in comparison with 82 per cent of supervisors, 55 per cent of P-6 officials, and 38 per cent of P-7 and P-8 administrators. As explained above, much of these differences is attributable to increasing age as we go from district ranger to P-7 and P-8 administrator, but we suspect not all.

Forest service men have in general the interests of skilled tradesmen (Group IV), particularly farmers; of production managers; of engineers; and of public administrators. Few of them have the interests of scientists (Groups I and II), office workers (Group VIII), salesmen (Group IX), lawyers-writers (Group X), and men engaged in social service (Group V).

Commenting upon Table 2, a former administrator of the service writes: "It explodes the old fiction that Forest Service men are scientists. True, some of us have had a little scientific training and some perhaps is a good thing, but we attempt to recruit scientists as rangers. Those we get go in one of three directions: some quit, some get transferred to research work, and some become frustrated and are no good to themselves or the Service."

There are some striking differences between the interests of district rangers near the bottom of the organization and the administrators at the top. Such differences are indicated fairly well in Table 2, especially in the case of those occupational

TABLE 2

*Percentage of Forest Service Men Rated A and B+ on Occupational Interests
(i.e., Percentage Scoring 40 and Above)**

Group	Occupation	49 District rangers	44 Supervisors	20 P-6 Washington & Regional	16 P-7—P-8 Washington & Regional
III	Production Mgr.	64	55	55	57
IV	Mechanical Activities				
	Forest Service	88	82	55	38
	Farmer	86	59	50	6
	Carpenter	35	11	5	0
	Aviator	36	27	35	6
	Printer	28	13	15	0
	Policeman	34	23	35	6
	Math Sci Teacher	34	21	25	6
II	Physical Sciences				
	Engineer	36	32	40	38
	Chemist	24	12	35	31
	Mathematician	0	0	5	6
VIII	Office Activities				
	Office Work	26	16	30	13
	Banker	18	28	20	12
	Accountant	10	13	15	13
	Purchasing Agent	22	29	20	25
IX	Sales Activities				
	Realtor	22	29	10	0
	Sales Mgr	6	25	20	0
	Life Insurance	8	11	15	6
XI	President	16	18	20	32
V	Social Service				
	Public Administrator	38	66	85	100
	Personnel Mgr	14	23	50	38
	Y Physical Dir.	14	4	15	6
	Social Sci Teacher	8	11	30	19
	City School Supt.	2	11	25	19
	Y.M.C.A. Secy	4	7	31	12
	Minister	0	5	5	6
X	Linguistic Activities				
	Author-Journalist	8	14	10	31
	Lawyer	6	25	5	44
	Advertising Man	4	12	5	19
VI	Musician	6	0	0	0
I	Biological Sciences				
	Physician	16	14	15	25
	Architect	8	2	10	25
	Dentist	10	9	5	0
	Artist	2	2	10	6
	Psychologist	0	0	10	12
VII	Certified Public Accountant	0	2	0	6

* The 49 rangers are a selection from 190 cases, so selected that the mean forest interest score and standard deviation of the 49 cases are practically the same as for the 190 cases. The 44 supervisors were similarly selected from 100 cases. The 20 P-6 and 16 P-7 and P-8 administrators are all the cases in our possession. The P-6 group contains 5 men stationed at Washington and 15 at regional offices. The P-7 and P-8 group contains 5 regional foresters and 11 men stationed at Washington.

interests in which a fair number of forest service men are interested. Such differences are, however, better shown by mean scores, for such take into account the men who score low as well as those who score high. Differences of four⁶ or more in mean scores between district rangers and P-7—P-8 administrators are given in Tables 3 and 4, the former giving the occupational interests which decrease and the latter the occupational interests which increase, as one goes from district ranger to administrator. (In nearly every case the mean scores of supervisors and P-6 administrators lie very close to a plotted line connecting rangers and P-7—P-8 administrators. This fact adds support to the conclusions drawn from the data of district rangers and P-7—P-8 administrators alone.)

The interests which decrease (Table 3) are for the most part typical of mechanical activities, whereas the interests which increase (Table 4) are much more varied, being associated with administrative work, law-journalism, and social work.

The differences in interest scores in these two tables indicate that administrators differ in their interests from district rangers and, to a lesser degree, from supervisors. The differences imply that administrators are selected on a different basis from that used in the original selection of district rangers. This relationship will be found in most organizations, for administrators differ from the rank and file both in abilities and in interests.

One of the most notable differences in interests between district rangers and P-7—P-8 administrators is in the interests of public administrators. Only 38 per cent of district rangers rate A and B+ compared with 66 per cent of supervisors, 85 per cent of P-6 administrators and 100 per cent of P-7—P-8 administrators.

⁶ Differences of 7 and 8 are statistically significant, i.e., have critical ratios of 3.0 and over, judging from a number of calculations, for example

		Forest Service		Farmer		Public Administrator		Personnel Manager	
		Diff	C.R.	Diff	C.R.	Diff	C.R.	Diff	C.R.
Ranger	vs Super	5.0	2.3	5.0	2.8	4.5	2.4	5.5	2.4
"	P-6	8.0	2.9	7.0	3.0	11.5	4.8	12.5	4.3
"	P-7—P-8	16.0	6.1	14.5	7.6	12.0	5.9	11.0	3.7
Super	vs P-7—P-8	12.5	4.8	9.5	4.7	7.0	3.4	5.5	1.9

TABLE 3

Occupational Interests in which Rangers Score at Least 4 More than P-7—P-8 Administrators, also Differences in Scores of Younger and Older Rangers and Supervisors

	District rangers*	P-7—P-8	Difference	Difference in scores of 30- and 50- year-old men	
				Rangers	Supervisors
Forest Service	52.5	36.5	-16.0	-10.3	2.0
Farmer	47.5	33.0	-14.5	2.3	2.5
Aviator	38.0	25.0	-13.0	-7.0	-1.0
Policeman	37.5	23.5	-14.0	7.5	4.0
Carpenter	36.0	14.5	-21.5	7.7	9.5
Printer	35.5	24.0	-11.5	.7	5.0
Math. Sci. Teacher . . .	35.0	29.0	-6.0	-5	3.5
Purchasing Agent . . .	33.5	29.5	-4.0	2.5	-5
Realtor	33.5	29.5	-4.0	2.7	-1.0
Office Worker	31.5	26.5	-5.0	3.7	3.5
Dentist	30.0	21.5	-8.5	1.1	2.0
Y. Physical Director . .	27.5	23.0	-4.5	-3.2	-5

* The rank order of occupations based on A & B+ ratings of rangers in Table 2 agrees very closely with rank order based on mean scores of rangers, some of which are given in this table and in Table 4. (The correlation between the two is .976.)

The public administrator scale is new. It is based on the interests of 518 men engaged in public administration. Included in the group are 46 supervisors and administrators of the forest service. The data in Table 2 are based on these 46

TABLE 4

Occupational Interests in which Rangers Score at Least 4 Less than P-7—P-8 Administrators; also Differences in Scores of Younger and Older Rangers and Supervisors

	District rangers	P-7—P-8	Difference	Differences in scores of 30- and 50- year-old men	
				Rangers	Supervisors
Public Administrator . .	39.5	51.5	12.0	-11.0	-4.0
President	30.5	35.5	5.0	.7	-5.0
Personnel Manager . . .	28.0	39.0	11.0	-13.8	-4.0
Author-Journalist . . .	27.0	34.0	7.0	-2.8	-3.0
Lawyer	25.5	37.5	12.0	-5.5	-2.0
Advertising Man	25.0	33.0	8.0	-4.0	-5.5
City School Supt. . . .	21.5	34.0	12.5	-7.3	-5
Y.M.C.A. Secy	21.5	25.5	4.0	-2.3	2.5
Mathematician	20.5	25.5	5.0	-3.3	.5
Minister	16.0	22.5	6.5	-4.8	3.5
C.P.A.	18.0	30.0	12.0	-7.3	-6.5
Psychologist	13.5	25.5	12.0	-9.8	-3.0

men and in addition 34 supervisors-administrators and 49 rangers not included in the public administrator criterion group.

Scores on the public administrator scale correlate highest with scores on the scales of personnel manager (.75), city school administrator (.55), Y.M.C.A. secretary (.53), social science teacher (.53) and Y.M.C.A. physical director (.52). Since forest service administrators rate so very much higher than rangers on the public administrator scale it is to be expected that they will score higher on the social service occupations. The data in Table 2 do not show such increases except in the case of personnel manager. This is true because, with the exception of personnel manager, few forest service men score high on these occupational interest scales. Examination of mean scores shows, however, a steady rise in these interests from ranger to the P-7—P-8 level (see Table 4). But it must be emphasized that at the top level the mean score is 39 on the personnel manager scale, 34 on the city school administrator scale, 29 on the social science teacher scale, and 25 to 22 on the other three social service scales. Forest service men on the whole score low on the social service scales, scales which are related significantly to the interests of public administrators. This is true even of the forest service administrators.

Younger Men More Similar to P-7—P-8 Administrators than Older Men

The fourth column in Tables 3 and 4 gives the difference in mean scores of 30(25-34)- and 50(45-54)-year-old rangers. Most of the differences are reversed from those between district rangers and P-7—P-8 administrators. That is, the older rangers differ from top administrators more than younger rangers. The same conclusion applies equally well to younger and older supervisors (last column in Tables 3 and 4).

The following correlations between interest profiles of groups of forest service men tell the same story.

30- vs 50-year-old rangers	861
30- vs 50-year-old supervisors	888
30-year-old rangers vs P-7—P-8	402
50-year-old rangers vs P-7—P-8	105
30-year-old supervisors vs P-7—P-8	769
50-year-old supervisors vs. P-7—P-8	569
P-6 vs P-7—P-8	578

The interests of younger and older rangers are quite similar; the same is true of supervisors. Rangers' interests are not very similar to the interests of P-7—P-8 administrators, but the interests of the younger rangers are somewhat more similar than those of the older rangers. Supervisors' interests are much more akin to the interests of top administrators. Particularly is this true of the younger supervisors, whose interests are more similar to those of P-7—P-8 men than are the interests of P-6 administrators.⁶

The differences in interests of district rangers and administrators suggest that the man who is most typically a ranger is not likely to rise above the rank of supervisor and that promotions above the rank of supervisor are in terms of interests which are possessed by only a minority of district rangers. One former administrator comments, "Time after time I have seen our top rangers promoted only to lose interest and become mediocre, or at least no longer outstanding." A very real problem here as elsewhere is "how to determine in advance who will respond to promotion and who will not."

What we know about interests indicates that they are fairly permanent, especially among adults. There are cases where they have changed appreciably but such appear to be exceptions to the rule. If we employ the data on younger and older district rangers and supervisors as indicative of the changes attributable to increasing age, then such changes are in the wrong direction—older men are less like administrators than are younger men. If we assume that interests are fairly permanent and that the above changes are not attributable to increasing age, then it would appear, as suggested above, that some older men with distinctly ranger but not administrative interests are dropping out of the service when promotions are not forthcoming.⁷

⁶ The difference of .297 between the two correlations of different age groups of district rangers has a critical ratio of 1.4, and the difference of .200 between the two groups of supervisors has a critical ratio of 2.3. Neither of these differences is statistically significant.

⁷ The data suggest that district rangers are well selected for that work. But they do not possess the interests characteristic of administrators. Their pay is small considering their responsibilities. They should be rewarded by increased status and pay, not through promotion into a different type of work, but by keeping them on their present work for which they are suited and which they enjoy.

One cannot help wondering if the forest service is recruiting at the bottom enough men typical of top administrators to provide a good assortment from which in later years to select the leaders of the organization.

Recreational Interests

In early days the forest service was concerned with the management of forest land involving lumbering, grazing, the construction of roads and buildings, and the fighting of fires. Then the public discovered the forests were a wonderful place for a vacation. This has focused greater emphasis upon the activity of handling people. Under the circumstances it is natural to ask, do the men selected for the original purposes of the service also possess the interests of men dealing with people? Is there any evidence that the younger men who have been selected in recent years have more social interests than the older men?

If we postulate that the occupations in Group V (see Table 2) typify the men who "handle others for their presumed good," then we can measure the extent to which forest service men possess social interests by noting their scores on the occupations in this group. Reference to Table 2 makes clear that few forest service men have such interests. The percentages are low for five of the six occupations and not at all high in the case of the sixth, i.e., personnel manager.

TABLE 5

Interests of Recreational Administrators, also Differences in Scores of Recreation Men and Forest Service Men

Occupational interests	Mean score of recreation administrator	Differences in score between recreation men and			
		District rangers	Superv	P-6	P-7-P-8
Public Administrator ..	48	- 8	- 4	3	4
Personnel	44	-16	-10	- 3	- 5
Social Science Teacher .	44	-17	-16	-12	-14
City School Supt . .	42	-20	-15	-12	- 8
Y.M.C.A. Secy . . .	41	-19	-16	-12	-15
Y. Physical Director .	40	-12	-12	- 9	-17
Lawyer	36	-10	- 1	- 5	2
Math-Science Teacher ..	35	0	- 4	0	- 6
Minister	35	-19	-17	-12	-12
Average deviation . .	.	-13 4	-10 5	- 7.5	- 9.2

If we postulate that the interests of public administrators engaged in recreational work typify the social interests which forest service men should now possess, we may then employ the interests of recreational men as a standard against which to check forest service interests. The nine occupational interests on which recreation administrators score 35 and higher are given in Table 5. It will be seen that six of the nine constitute the occupations in Group V referred to above. It will also be recalled that these interests correlate significantly with the interests of public administrators in general.

Forest service men score low on most of the interests listed in Table 5 (differences of 8 are statistically significant in almost every case). There is better agreement between the interests of recreation administrators and forest service men as we go from ranger to P-7—P-8 administrators, with the exception that P-6 administrators are slightly more similar to recreation administrators than are P-7—P-8 administrators.

When all 36 occupational interests are taken into account instead of only nine, as in Table 5, we have the following correlations between the interest profiles of recreation administrators and sub-groups of forest service men:

Recreation vs	30-year-old district rangers	- 01
" "	50-year-old district rangers	- 06
" "	30-year-old supervisors34
" "	50-year-old supervisors31
" "	P-651
" "	P-7—P-835

Evidently there is no relationship between the interests of recreation administrators and the interests of rangers, only a slight relationship in the case of supervisors and P-7—P-8 administrators, and some relationship in the case of P-6 administrators.

The above correlations between the interests of recreation administrators and forest service men may be compared with the following correlations between recreation administrators and eight other groups of public administrators:

Recreation vs	Personnel men in public service80
" "	Social insurance administrators75
" "	Welfare administrators86
" "	Publicity men29
" "	Statistician16
" "	Public health officials	-.12
" "	Engineers	-.25
" "	Chemists-Physicists	-.42

Seemingly, if the forest service is to handle the problem of recreation within the forests it must have men in the organization who understand such problems and genuinely enjoy dealing with them. Such interests are different from the interests of the typical forest man. Two ways of meeting the situation occur to us. First, men who possess both types of interest might be brought into the service. Second, men who are possessed of the recreational type of interest might be brought into the service to be specialists in this field. The former is a doubtful procedure because there are not many men who possess both sets of interests. (Forest service interests correlate with the interests of the occupations listed in Table 5 as follows. Public administrator, .21; Personnel manager, -.01, social science teacher, -.13; city school superintendent, -.25; Y.M.C.A. secretary, -.07; Y.M.C.A. physical director, .39; Lawyer, -.61; Math-Science teacher .68; and Minister, .00.)

The second procedure would force a rearrangement by which recreational activities would at least be directed, if not carried on, by specialists. This would not be so convenient as the present procedure of having rangers carry on all types of activities. Whatever the organization, it certainly appears that there are not enough forest service men with the interests of recreation men to carry on such work enthusiastically.

Since the interests of administrators directing public recreational work correlate significantly with the interests of administrators in general, it is possible that adding such men to the forest service might result in increasing the number of younger men who would be selected later on for administrative work in the forest service.

Conclusion

This report is restricted to the interests of forest service personnel. Abilities as well as interests must be considered before complete answers to the questions raised here can be obtained. The data suggest:

1. Forest service men have in general the interests of skilled tradesmen, particularly farmers; of production managers; of engineers; and of public administrators. Few of them have the

interests of scientists (Groups I and II), office workers (Group VIII); salesmen (Group IX), lawyers-writers (Group X), and men engaged in social service (Group V).

2. The older the man in the forest service, regardless of whether he is a district ranger or top administrator, the less he has the interests of district rangers-supervisors. Such decrease in interest score is apparently not attributable to increasing age, or shift in type of man entering the service, but to men with high scores leaving the service in greater numbers than men with low scores. Another bit of evidence in support of this explanation is that younger district rangers have interests somewhat more similar to those of top administrators than do older district rangers, the same thing being also true of supervisors.

3. District rangers differ in their interests from administrators in the forest service. The former have stronger mechanical interests and weaker interests associated with law-journalism, social work, and administrative work typified by city school superintendent, personnel managers, president of a business concern, and public administrator.

One cannot help but wonder if there are enough younger men in the service with interests of administrators so that the higher positions can be well filled twenty to thirty years hence.

4. Relatively few of the forest service personnel have the interests found among recreation administrators. But in recent years the forests have come to be used by millions of people for recreational purposes.

Should not some forest service personnel be specifically selected to handle the recreational facilities of the service instead of expecting this function to be taken care of by men already heavily loaded with other work and not possessing the interests of recreational enthusiasts?

5. It is likely that the rank and file of employees in most organizations differ in their interests from the executives.

Should men be selected to fill the lower positions with the hope that some will eventually fill the few higher positions as vacancies occur, or should some men be definitely selected for future placement in higher positions?

A GROUP TESTING PROGRAM FOR THE MODERN SCHOOL

WARREN G FINDLEY

New York State Education Department

IT is the fundamental thesis of this paper that the modern school has certain definite functions and characteristics which, in the light of the findings of educational psychology, call for a group testing program having equally definite characteristics.

Characteristics of the Modern School

The modern school is characterized, first of all, by being charged with responsibility for educating all the children of all the people from school entrance until completion of grade 12 at age 18. At present some pupils do not complete grade 12 or remain in school until 18 years old, but increasing proportions do so, curriculum planning is based on this premise, and research is being pointed in the direction of finding why the schools fail to hold those that now leave without graduating. This trend toward the twelve-year common school has resulted in increasing individual differences in the pupil population in the upper grades with respect to interests and life goals, if not also with respect to intellectual ability.

A second important characteristic of the modern school, especially in the elementary grades, is the tendency to promote regularly practically all pupils. This is based on the psychological finding that the individual differences most significant in total development are social factors associated more closely with chronological age than with intellectual achievement. The practice of regular promotions finds further justification in the readily observed and experimentally tested deleterious effects of retardation on the morale of the retarded pupil. This trend in promotions policy has produced increased indi-

vidual differences in intellectual development in each school grade, so that teachers have been encouraged to "take each child where he is" and attempt to further his development as much as possible by adaptation of instruction to individual pupil needs.

A third characteristic of the modern school that bears on our problem has grown rapidly under the pressure of war findings that many high-school graduates have failed to maintain skill and ability at levels attained earlier. This condition has been found especially true of arithmetic. As a result the secondary school is now taking responsibility for maintaining and developing skills and abilities in basic areas like reading, arithmetic, and geography, previously accepted as mastered in the elementary school.

A fourth significant characteristic of the modern school is the emphasis on following and guiding total personality development of individual pupils cumulatively and constructively, and in line with this bringing to parents a clearer and less formal report of their children's progress and mastery of the work of the school. The old report card is giving way to interpretative communications, often including other matters besides scholastic achievement, calculated to provide the basis for better understanding and cooperation between the school and the home.

Other characteristics of the modern school might be noted—many of them significant—but the above will suffice for the problem under discussion.

Implications for the Testing Program

The twelve-year common school—more than twelve years insofar as kindergarten and nursery school are added at the beginning—and the corollary acceptance by the secondary school of responsibility for maintaining and extending development of basic skills have made it possible to relax the attitude of frequent and grim testing in the early grades. There is no longer the necessity to view each year's work as possibly the last. With extended universal schooling it becomes possible to plan extended programs of instruction looking toward a

final pointing up at the end in more comprehensive examinations of the outcomes of the whole program. The best achievement at the end of the twelve years becomes the criterion and the justification for practices that lead to such success. In the pupil's whole school career the effort of the school properly goes to studying, guiding, and furthering his development toward ultimate achievement.

A program of testing progress along the main courses of the curriculum is called for if progress is to be guided. Such a program in New York State is called just that, a program of "progress testing" and "progress tests." *Reading Progress Tests* are designed to reflect progress in reading comprehension from grade 4 through grade 12. *Mathematics Progress Tests*, just issued, are similarly to reflect progress in mastering general mathematics from grade 4 to as high a grade as may involve general mathematical instruction. Similar tests are planned for the understandings, skills and appreciations in written English, science, health and safety, social studies, study skills, appreciation of literature, art appreciation, and music appreciation.

A further feature of this testing program derives from the tendency toward uniform promotions and the emphasis on interpreting individual achievement both within the school and to the parents. In such a program, testing must reflect progress along specific and directly improvable lines rather than in general terms of subjects. The *Reading Progress Tests* measure progress in three identifiable aspects of reading comprehension: ability to obtain detailed understanding, ability to discern central thoughts in passages, and ability to recognize meanings of words. The *Mathematics Progress Tests* measure progress in five aspects of problem solving: social information and quantitative concepts that are the basis for understanding the context of mathematical situations or problems, computational skill, ability to choose the operation or operations required to solve particular problems, ability to determine whether one has sufficient data to solve problems and to reject data irrelevant to the solutions, and finally ability to carry through the complete act of problem solving, using all the abilities separately measured in the other parts of the tests.

Many standardized tests of commercial publishers covering reading comprehension and English usage involve similar subparts devoted to particular skills in those areas. The *Iowa Every-Pupil Tests of Basic Skills*, especially *Test B: Work-Study Skills*, are so organized. *Test B* measures progress along five lines: reading maps, knowing sources of information, using the dictionary, using an index, and reading charts, graphs and tables.¹

Tests thus subdivided provide the teacher with information about the relative strengths and weaknesses of the class as a whole and of individual pupils that permits planning of instruction to meet demonstrated needs. Summaries of achievement in this detail help administrators obtain more real understanding of the progress and mastery being accomplished under their general supervision. Profile charts showing the achievement of individuals and classes in the various skills bring out strengths and weaknesses especially clearly to both teacher and administrator. They also motivate pupil effort toward progress and provide a basis for intelligent understanding by parents of their children's status and progress. Under proper conditions, pupils may acquire satisfaction and hence motivation in plotting their own profiles from year to year in the areas tested.

A long step forward can be taken by establishing an annual fall testing program of progress tests. If promotions policy is to result in wide individual differences in every grade and promotion into a grade is no longer to be based merely on meeting minimum standards of achievement, fall testing in basic skills and areas of understanding will provide data essential to the teacher as she begins work with her new class. Testing at this

¹ The standardized testing approach described in preceding paragraphs applies well to testing progress in the mastery of skills. It has its limitations when applied to the content aspects of the social studies and science. Tests of content in these areas need to be adapted to varying local situations and programs and need to have a timeliness that standardized materials cannot maintain. It is true that much of the content of these areas that is important remains the same from year to year and from decade to decade. Standardized tests of these aspects, however, necessarily place a premium on permanent content at the expense of the timely and the local. Annually prepared school-wide, city-wide, or district-wide tests of the aspects neglected in standardized tests are essential if standardized tests of content are to be kept from exerting a restrictive influence on instruction through presenting a biased evaluation of content outcomes.

time gives up-to-the-minute evidence on what learning has survived the summer vacation to become the basis for further progress. For this very reason, such evidence is also more relevant to the administrator's interest in appraising the status and progress of his charges, than that which has commonly been obtained at the end of the school year.

Fall testing avoids the undesirable practice of cramming. Teachers will be less inclined to teach for specific tests. Testing at this time places a wholesome emphasis on collaboration between teacher and pupil to achieve goals of instruction in the year ahead.

If testing in the basic areas is done each year at the opening of school in the fall and the results are entered on cumulative records kept for each pupil, each teacher is provided not only with data concerning the current year's testing, but also with evidence of the previous progress of her new pupils. Achievement on this year's test takes on added significance when judged in the light of previous achievement and relevant notations about health, social, and emotional development, etc., that may be entered on such records. A pupil's seemingly mediocre achievement this year may represent significant improvement over his very poor achievement of earlier years, as a result of serious effort on his part and, perhaps, special assistance by his teacher. Such progress, which could not be inferred from the results of this year's testing alone, merits attention and encouragement.

What has been said above applies generally to the group testing of pupils from grade 4 through grade 12. In the primary grades, informal testing is generally to be preferred to standardized testing. In these grades children are still acquiring elementary skill in reading and the minimum vocabulary essential to dealing with standardized test materials. Printed standardized tests call for reading to such an extent as many times to make reading ability the main factor actually tested, regardless of the title and content of the particular tests, or they are devoted to testing only the mechanics of subjects, like computing in arithmetic or spelling in English. Until children have matured sufficiently to handle well-rounded tests, it

is better to omit formal group testing of the mechanical aspects of subjects. Until it is practicable in actual test situations to relate computation to problem-solving and spelling to effective writing, it is better not to risk drawing undue attention and effort to those aspects of the subjects that are easiest to test.

There are, however, some standardized tests that are useful in these grades. In the first grade or at the end of kindergarten, "readiness" tests may be helpful to teachers in judging pupils' readiness to undertake beginning study of the printed word and other work of the first grade. At the beginning of second and third grades carefully selected standardized reading tests will aid many teachers to appraise the general level of reading ability of each of their pupils, as in later grades the more detailed program of progress testing provides data on many counts.

A word regarding intelligence testing. Group intelligence testing—and group testing is most common—will add little to what is learned from a reading readiness test at primary levels. The unreliability of intelligence measures from group testing at these levels makes it wise to avoid such testing and the temptation the testing brings to enter a very doubtful I.Q. on the pupil's early record. In the intermediate grades the situation is quite different. Pupils are generally adaptable to group testing and results of intelligence testing provide a helpful clue in determining the approach to be made to a pupil having difficulty in his studies. One with a high I.Q. may be presumed to have prospects of considerable improvement if the specific source of difficulty in learning can be found, one with a low I.Q. may be expected to have considerable difficulty at least in the immediate future in most of his learning and should therefore be encouraged in his studies, but not exhorted to seek to attain tremendous progress. Annual administration of a group intelligence test at the beginning of grades 4, 5 and 6 should yield data immediately useful. The three separate testings will result in establishing a reliable measure for future as well as immediate reference. At the junior and senior high-school levels it may again be questioned whether anything useful in general school practice is accomplished by intelligence testing. What

has been learned of I.Q. in the intermediate grades and what may be ascertained by fall testing of skills in the upper grades provide all the necessary evidence of status and progress and a sufficient basis for prognosis of general or special success in advanced grades.

By way of summary, it may be said that the characteristics of the program of group testing recommended for the modern school are:

- 1 A twelve-grade testing program corresponding to the twelve-grade instructional program, with achievement by the end of grade 12 the ultimate measure of effectiveness of the total program.

- 2 Continuous testing of progress throughout the pupil's school career, recorded on cumulative records, as the basis for guiding his development toward ultimate achievement.

3. Use of tests yielding significant part scores related to specific, improvable skills, so that instruction may be related to demonstrated needs of individuals and whole classes.

- 4 Annual fall testing in each grade so that up-to-the-minute data may be available to all involved in guiding pupil development—teachers, administrators, parents, and the pupils themselves—at a time that is helpful and especially related to looking ahead.

5. Spring use of formal tests below grade 4.

6. Use of group intelligence tests annually in the intermediate grades (grades 4 through 6).

Nothing has been said of diagnostic testing of a refined sort. Such testing is best conducted on an individualized basis, possibly following up leads suggested by the group testing program, but not as a part of the group testing program.

Group testing of interests and attitudes is not treated here. A group testing program in these areas is well justified, but falls outside the limits of this brief paper.

REPLIES OF PSYCHOLOGISTS TO SEVERAL QUESTIONS ON THE PRACTICAL VALUE OF INTELLIGENCE TESTS

ARTHUR KORNHAUSER

Bureau of Applied Social Research, Columbia University

AN earlier report summarized the answers of a panel of mental test specialists to one part of a questionnaire on tests.¹ The present paper deals with the less technical questions which were asked at the same time.² Both reports are based on the replies of 79 psychologists. The questions were originally sent to 85 persons in this field who were chosen as representative of the most competent "experts" on mental measurement. The selection was based upon the pooled judgments of 6 advisory specialists who were consulted for the purpose. The names of the psychologists who participated in the poll are listed in connection with the report in the preceding number of this journal.³

Question 1

In your judgment, how well do intelligence tests meet the practical needs for classifying people as to general mental ability in the Army, in schools, and in industry?

	<i>In the Army</i>	<i>In schools</i>	<i>In business and industry</i>
Extremely well, with a very small amount of error .	7%	19%	7%
Rather well, much better than is done without tests	81	78	60
Not very well, but somewhat better than without tests	12	3	33
Not at all well; little or no better than without tests .	0	0	0
	<u>100%</u>	<u>100%</u>	<u>100%</u>
Number of cases	76	77	76
No answer or not classifiable	3	2	3

¹ This journal, Spring issue, 1945, pp. 3-15

² The answers to these questions were obtained for the purpose of giving a popular summary of expert views on the matter to the public. This popularized report is published in *The American Magazine* for July 1945, as part of a new monthly "Poll of Experts" project.

³ Loc cit Footnote p. 3

An interesting result appears when the experts are divided according to whether or not they state that they have done test work in the armed forces, in schools, and in industry. The tests are rated a little higher in each field of application by those experts who have *not* worked in that particular field. While the relationships are slight and the numbers small, a consistent trend is present. The figures are as follows:

	Has the expert done test work in the armed forces?	
	<i>Has</i>	<i>Has not</i>
The tests do "extremely well" in the Army	2%	13%
The tests do "not very well" in the Army	14%	10%
	(n=43)	(n=31)
	Has the expert done test work in schools?	
	<i>Has</i>	<i>Has not</i>
The tests do "extremely well" in schools	19%	25%
The tests do "not very well" in schools	3%	0%
	(n=59)	(n=16)
	Has the expert done test work in business?	
	<i>Has</i>	<i>Has not</i>
The tests do "extremely well" in business	0%	9%
The tests do "not very well" in business	38%	32%
	(n=21)	(n=53)

Similar comparisons by age of the respondents show no tendency for the younger and older to differ in their ratings of test accomplishment. The same is true of clinical compared with non-clinical psychologists (classified according to their report of their own principal types of work).

When those who indicate "psychometrics" as a field of work are compared with others, a slight tendency is observed for the psychometricians to use the high and low rating categories a little more than do the others. Thus, the percentages of psychometricians saying "extremely well" with respect to the value of tests in the Army, in schools and in industry respectively are 9, 27, and 9; for non-psychometricians the corresponding figures are 3, 9, and 3. (The perfect 3 to 1 pattern is fortuitous!) For the low ratings of test accomplishment, the 2 sets of percentages are these: Psychometricians 14, 2, and 37; others 9, 3, and 27. This tendency may imply that the psychometricians had greater confidence in their own evaluation of test accomplishment than did the other psychologists.

Illustrative Comments of the Respondents on Question 1

Test administration in schools is usually slightly more careful than in the Army and in business and industry and consequently the results are slightly more reliable.

Better in Army and in business and industry because of greater heterogeneity of populations dealt with—less well in schools because of greater homogeneity. On the other hand, “general mental ability” is of more importance in schools and if adequate testing is done the job can be done extremely well in schools.

I would place the usefulness of a general classification test to schools and the Army above that to industry. This is because the problems in the last field are very frequently highly specific. Despite the greater complexity of Army problems I consider usefulness to the Army almost equal to that in the schools on the ground that tests classify the men much more accurately than other techniques under conditions where little time is available for processing.

The fields are still too general. For example in my present work in a large aircraft plant I find intelligence tests meet the practical needs for classifying accountants better than they do foremen and sub-foremen in the factory. Same would be true, I imagine, in the case of the Army.

Group tests below age 9 or grade 4, even when carefully administered, frequently yield unreliable determinations. At higher levels such tests do rather well in school.

The tests measure a type of general mental ability. The question is how important this rather abstract type of ability is in practical affairs. In school the importance is fairly clear, but in the Army or in business and industry the importance seems more questionable, and depends upon the specific job being studied.

The practical need is to classify people in terms of several specific *types* of mental ability, not in terms of *general* mental ability.

The concept of general mental ability has been largely discarded by mental test *experts* with experience in practical testing work outside of the school situation.

(Finally, there is one very long and critical reply from which two paragraphs are quoted, as follows:) Question 1 presupposes that there is some trait which one may properly call “general mental ability,” and Question 2 likewise presupposes that there is a trait which one may properly call “mental ability,” both traits being (a) unequivocally defined and (b) capable of being detected and measured *independently of the test that is employed to test them*. Are there such traits? If so, what operations determine them? If they exist, and if you define the appropriate operations, then both questions

have meaning. Only, the answer is to be found in an actuarial or contingency-table. It needs no census of experts' beliefs about what the proper answer would be. If there are no such traits as "general mental ability," or plain "mental ability," determinable independently of the tests which you have in mind, then every answer that you provide for will be factually false, since it presupposes a false assumption. . . .

If the tests are to be used, not to classify people as to "general mental ability" but rather to predict success in training, or the like, the answer is to be derived from the coefficient of validity, not by pooling personal impressions. Speaking by and large, these coefficients have run between (say) .3 and .5. For group prediction and rough screening they are valuable; for individual prediction almost worthless.⁴

Question 2

(a) How dependably do intelligence tests (the usual group tests for adults) measure the mental ability of *the individual person*? How safely can a person accept his test rating as a correct indication of where he will continue to stand in mental ability relative to others?

(b) What is your answer to the same question concerning the mental test rating of *the individual school child*?

	Ques. (a)	Ques. (b)
Dependable measure of individual's ability; can be relied upon	8%	7%
Moderately dependable, seldom far wrong . . .	74	76
Doubtfully dependable; often in error; not to be accepted without confirmation	18	17
Not at all dependable, likely to be misleading, should not be taken seriously	0	0
	<hr/> 100%	<hr/> 100%
Number of cases	74	75
No answer or not classifiable	5	4

⁴A comment or two may be appropriate in defense of our questions about "intelligence tests" and measuring "general mental ability," in reply to these last 3 quotations (4 others expressed similar but less definite concern). The principal answer is that these concepts *are* widely used, both by psychologists and others, and that an immense amount of mental testing is aimed at measuring this general intellectual ability. Perhaps it would be better, as one respondent suggests, to employ the term "average mental ability" but prevalent usage seemed to warrant the assumption that test experts would recognize what is referred to in questions about "intelligence tests" and the practical classification of people on the basis of these measurements. There can be no doubt that practical efforts are frequently directed at predicting general alertness, adaptability, potentialities for learning, etc. The private convictions of particular psychologists regarding the futility of trying to measure these qualities is scarcely reason enough to justify discarding the concept and refusing to ask questions which pertain to it.

The other general criticism which is contained in the final quotation, to the effect that questions of the type we have asked can be answered simply by stating the size of correlation coefficients and standard errors, likewise seems untenable. There are many confusing statistical results on these matters which make it necessary for the "experts" to judge what are the typical or representative statistical findings. Moreover, even after particular figures are accepted, a problem remains as to the justifiable interpretation of the figures for practical purposes. It is precisely these final conclusions based upon the coefficients (plus the less quantitative evidence) which we were trying to ascertain.

Replies to these questions were compared for clinical and non-clinical psychologists and for psychometricians versus non-psychometricians. The clinical and the non-psychometric respondents are decidedly more harsh in their judgments regarding the dependability of the tests. On Question 2a the "doubtfully dependable" category (the lowest used by any rater) contained 31% of the clinical psychologists' answers and 39% of the non-psychometricians' as against 10% and 5% for the contrasting groups. Question 2b showed similar but less marked differences: 21% and 26% compared with 12% and 9%. On both questions the top rating category reveals slight differences of the same kind—that is, the psychometricians and non-clinicians tend to be a little more favorable in their estimates.

It seems probable, judging from some of the comments by the respondents, that the explanation for these differences may lie in the fact that a number of the measurement psychologists and non-clinicians answered in terms of simple short-run re-test reliability coefficients. The others tended to go beyond such figures and to consider also the varied and unequal conditions affecting individuals and their differing rates of growth over prolonged periods of time. The questions were intended to cover this larger problem rather than to refer to the narrower question of test reliability in a technical sense.

Illustrative Comments of Respondents on Questions 2a and 2b

Most group tests place the non-reader or slow reader at so great a disadvantage as to make results questionable until reading ability is determined.

But test results should be checked by reference to other sorts of evidence such as accomplishment in school and in occupation, and supplemented, when there are discrepancies or doubt, by tests administered individually.

Caution, of course, is necessary in the presence of unusually high or low ratings; verification with a more refined individual test is desirable in such instances.

Verbal tests are satisfactory for scholastic needs in school, but not for all other school needs. Likewise factors extraneous to test data may invalidate their (test data) classification value, e.g., low effort may negate high intelligence and vice versa. Wide diversity in values of different tests.

have meaning. Only, the answer is to be found in an actuarial or contingency-table. It needs no census of experts' beliefs about what the proper answer would be. If there are no such traits as "general mental ability," or plain "mental ability," determinable independently of the tests which you have in mind, then every answer that you provide for will be factually false, since it presupposes a false assumption. . . .

If the tests are to be used, not to classify people as to "general mental ability" but rather to predict success in training, or the like, the answer is to be derived from the coefficient of validity, not by pooling personal impressions. Speaking by and large, these coefficients have run between (say) .3 and .5. For group prediction and rough screening they are valuable; for individual prediction almost worthless.⁴

Question 2

(a) How dependably do intelligence tests (the usual group tests for adults) measure the mental ability of *the individual person*? How safely can a person accept his test rating as a correct indication of where he will continue to stand in mental ability relative to others?

(b) What is your answer to the same question concerning the mental test rating of *the individual school child*?

	Quer. (a)	Quer. (b)
Dependable measure of individual's ability; can be relied upon	8%	7%
Moderately dependable; seldom far wrong . . .	74	76
Doubtfully dependable; often in error; not to be accepted without confirmation	18	17
Not at all dependable, likely to be misleading, should not be taken seriously	0	0
	<hr/> 100%	<hr/> 100%
Number of cases	74	75
No answer or not classifiable	5	4

⁴ A comment or two may be appropriate in defense of our questions about "intelligence tests" and measuring "general mental ability," in reply to these last 3 quotations (4 others expressed similar but less definite concern). The principal answer is that these concepts *are* widely used, both by psychologists and others, and that an immense amount of mental testing is aimed at measuring this general intellectual ability. Perhaps it would be better, as one respondent suggests, to employ the term "average mental ability" but prevalent usage seemed to warrant the assumption that test experts would recognize what is referred to in questions about "intelligence tests" and the practical classification of people on the basis of these measurements. There can be no doubt that practical efforts are frequently directed at predicting general alertness, adaptability, potentialities for learning, etc. The private convictions of particular psychologists regarding the futility of trying to measure these qualities is scarcely reason enough to justify discarding the concept and refusing to ask questions which pertain to it.

The other general criticism which is contained in the final quotation, to the effect that questions of the type we have asked can be answered simply by stating the size of correlation coefficients and standard errors, likewise seems untenable. There are many confusing statistical results on these matters which make it necessary for the "experts" to judge what are the typical or representative statistical findings. Moreover, even after particular figures are accepted, a problem remains as to the justifiable interpretation of the figures for practical purposes. It is precisely these final conclusions based upon the coefficients (plus the less quantitative evidence) which we were trying to ascertain.

Replies to these questions were compared for clinical and non-clinical psychologists and for psychometricians versus non-psychometricians. The clinical and the non-psychometric respondents are decidedly more harsh in their judgments regarding the dependability of the tests. On Question 2a the "doubtfully dependable" category (the lowest used by any rater) contained 31% of the clinical psychologists' answers and 39% of the non-psychometricians' as against 10% and 5% for the contrasting groups. Question 2b showed similar but less marked differences: 21% and 26% compared with 12% and 9%. On both questions the top rating category reveals slight differences of the same kind—that is, the psychometricians and non-clinicians tend to be a little more favorable in their estimates.

It seems probable, judging from some of the comments by the respondents, that the explanation for these differences may lie in the fact that a number of the measurement psychologists and non-clinicians answered in terms of simple short-run re-test reliability coefficients. The others tended to go beyond such figures and to consider also the varied and unequal conditions affecting individuals and their differing rates of growth over prolonged periods of time. The questions were intended to cover this larger problem rather than to refer to the narrower question of test reliability in a technical sense.

Illustrative Comments of Respondents on Questions 2a and 2b

Most group tests place the non-reader or slow reader at so great a disadvantage as to make results questionable until reading ability is determined.

But test results should be checked by reference to other sorts of evidence such as accomplishment in school and in occupation, and supplemented, when there are discrepancies or doubt, by tests administered individually.

Caution, of course, is necessary in the presence of unusually high or low ratings; verification with a more refined individual test is desirable in such instances.

Verbal tests are satisfactory for scholastic needs in school, but not for all other school needs. Likewise factors extraneous to test data may invalidate their (test data) classification value, e.g., low effort may negate high intelligence and vice versa. Wide diversity in values of different tests.

It would be better never to give anybody a single score called an "intelligence" score. Various aptitude scores should be used instead.

In my opinion incalculable damage has been done by the dogma of the "constancy of the I.Q." For example, in some school systems a child whose I.Q. at 9 years is 90 is not allowed to take algebra or a foreign language. Considering that the S.E. of prediction is about 9 points, one can see that some children are called dull at 9 who would be called normal or better at 11; while a very small proportion rated dull at one age would be rated near geniuses at another.

The use of group test results as predictors for individual cases is dangerous. High scores more meaningful than low, i.e., you deserve the score you get but maybe you should have more. Factors reducing scores (which are extraneous to the test itself) are far in excess of those raising scores.

In my estimation a good group intelligence test gives a moderately dependable estimate of the person's present functional level, but it does not predict his future standing as well as we like to think.

The above vote of confidence is predicated on the employment of the best tests; and one long enough to be reliable. Most short tests are not reliable.

My response here ("moderately dependable") refers to the "best" group tests. The widely used "self-administering" group tests are "doubtfully dependable."

The problem here is complicated by growth effects. A few children, destined to be normal, start growing late as if stunted. Repeated retesting, graphing the results, enables this to be detected.

I believe intelligence test ratings are a little more dependable for the older school child (high school) than they are for the younger ones.

Greater dependability with younger school children than with adults or older school children.

The younger the child, the greater the question.

Varies with the age of the child; very inaccurate in first years of life, and not dependable for adolescent years.

Whatever general mental ability exists is less subject to change in most cases after school-leaving than before.

I am more sure about this (dependability of person's test score) for school children than for adults because children are usually subjected to a similar school environment and abilities have a more equal chance to develop. With adults it is

possible to have differences occurring because of differences in environment—due to country of origin, work, etc.⁵

Question 2c

Are people with limited schooling rated correctly in comparison with persons who have more schooling, or do the former tend to be rated too low?

Limited schooling not a source of error	0%
Source of slight error	38
Source of considerable error	57
Source of serious error	5
	<hr/>
	100%
Number of cases	63
No answer or not classifiable	16

Here, too, comparisons have been made in terms of whether the respondent checked "clinical work" or "psychometrics" as a principal part of his activities. Again, the psychometricians give slightly more favorable evaluations than the non-psychometricians. They give the rating of "slight error" in 39% of the answers as compared with 31% for the non-psychometricians; "serious error" is given by 3% as against 10% by the others. More of the psychometricians proportionately refuse to answer the question as asked, however (22% as compared with 12%). Clinical psychologists use both the upper and lower extreme ratings a little more than do non-clinicians ("slight error" 44% versus 30%, "serious error" 12% versus 3%) and are more reluctant to answer (24% versus 13% give no answers).

Illustrative Comments of Respondents on Question 2c

The comments under this question most frequently noted that the answer turns on the kind of test used or the nature of the test content (i.e., the problem is more serious with group tests and verbal tests than with individual and non-verbal ones). Twenty-four of the respondents made this point. The only other ideas expressed in many answers were that the reply depends upon how *extreme* the educational lack; whether due to absence of educational *opportunity* or not; and what the

⁵ Interesting disagreements occur among these last six quotations. To some extent the apparent discrepancies doubtless stem from the ambiguity of the words "younger" and "older." Further inquiry would be required to find what other sources of disagreement are present.

nature of the cultural environment of the persons considered, apart from their schooling. Examples follow.

The verbal group tests place a considerable premium upon formal education, particularly on the first few grades. Non-verbal tests tend to reduce this effect. Recency of education is also an important factor.

Generally, the error is slight if exposure through at least the first six grades exists. (Another respondent says: Considerable error if less than 30 months of schooling.)

Again depends on type of test employed and how cautiously interpreted. Also depends on cultural level of environment within which limited schooling obtains. Also depends on possibly related language handicap not due to lack of schooling in own language.

This likelihood of error is particularly true for individuals who are dull. Unusually able individuals tend to rise above the limitations of schooling and to acquire such information by other means and from other sources.

Generally speaking, mental ability determines amount of schooling. Bright individuals with limited schooling make high scores, whereas dull individuals who have been kept in school a long time make low scores.

People with limited schooling rate lower than they will with more schooling. Whether this is an "error" depends on one's concept of intelligence.

The larger the adequacy of the education, the more accurately does the test reflect the true capacity of the individual.

In my opinion, the test tends to rate correctly those with limited schooling who have had ample opportunity for schooling; there is a source of considerable error where the group with limited schooling are those coming from communities where opportunities for schooling have been limited and where the limited schooling is the direct result of limited opportunities.

Source of slight error for typical American communities; about 85 to 90% of children. No general statement is adequate; for perhaps 5% of our school children the error may be serious, and "considerable" for another 10% (especially in certain regions).

Question 3

(a) Do you find any serious misunderstandings or false expectations about intelligence tests (or "I Q." tests) on the part of non-psychologists?

Do	92%
Do not	4
Doubtful or don't know . .	4
	<hr/>
	100%
Number of cases	78
No answer	1

(b) If you do, what are the principal misunderstandings?

The types of misunderstandings mentioned by the respondents have been classified as shown in the following list. The items are arranged in descending order according to frequency of mention. The percentages are based on the number of respondents who named any classifiable misunderstanding, namely, 74 persons. Since the parts of a single answer were often classified under several headings, the percentages total more than 100.

"Principal Misunderstandings"

Over-rating the test results; exaggerated belief in test validity, reliability, accuracy, constancy of I.Q., etc.	58%
Belief that a test measures all aspects of ability, neglect of separate abilities; use of I.Q. for purposes for which not intended	55%
Confusion in meanings of terms (I.Q., M.A., percentiles, intelligence, etc.), thinking of <i>any</i> test rating as an I.Q.; confusion of intelligence and information; wrong use of I.Q. applied to adults	46%
Tendency to go to extremes in appraising tests; they are wonderful or they are worthless	30%
Assumption that tests measure innate ability; that they are independent of environment	22%
Other misinterpretations of what the tests measure or of their limitations	19%
Failure to interpret scores in relation to norms or to think in comparative terms, misuse of norms	16%
Under-rating the test results; exaggerated disbelief in test validity, reliability, etc.	12%
Failure to recognize that some tests are better than others (group versus individual; limitations of verbal tests; etc.)	14%
Too much credence given a single measurement, regardless of how and where test was administered	10%

MEASUREMENT ABSTRACTS*

Altus, William D. "The Differential Validity and Difficulty of Subtests of the Wechsler Mental Ability Scale." *Psychological Bulletin*, XLII (1945), 238-249

This is a report of a study made at an Army Special Training Center to determine the validity of the Army Wechsler subtests for use in counseling, guidance and prediction. The subjects were illiterate army trainees, and the criterion was "success-failure" in attaining a degree of literacy, comparable to the fourth-grade public school level. English-speaking subjects were given the verbal subtests of the Wechsler scale, while non-English-speaking subjects were given the performance subtests. Biserual correlations were computed and significant differences obtained. "Arithmetic Comprehension" and "Similarities" were highest on the verbal section, while "Digit Symbol" and "Series Completion" were high on the performance scale. Francis Medland.

Bailey, H. W. and Dallenbach, K. M. "A Study of Selective Procedures and Educational Achievement of ASTP Trainees Processed by the STAR Unit at the University of Illinois." *American Journal of Psychology*, LVIII (1945), 1-24

The Specialized Training and Reassignment unit was organized to increase the efficiency of placement of candidates in Army Specialized Training Programs. Under its direction all candidates for ASTP took a basic battery of five tests. a) *Army General Classification* b) *Officer Candidate* c) *American Council on Education Psychological Examination* d) *Algebra* e) *Geometry*. These tests were measured against the "pass-fail" criterion for success in prediction. All showed "critical ratios" that were significant at the 5% level, the order of effectiveness of prediction from highest to lowest was. a) *Officer Candidate* b) *Algebra* c) *Geometry* d) *American Council on Education Psychological Examination* e) *Army General Classification*. Francis Medland.

Baldwin, E. F. and Smith, L. F. "The Performance of Adult Female Applicants for Factory Work in the Likert-Quasha Revision of the Minnesota Paper Form Board Test." *Journal of Applied Psychology*, XXVII (1944), 468-470

The data for the table of norms presented in this paper were secured from the scores made by 975 women tested at the Hawk-Eye Works of the Eastman Kodak Co., Rochester, N. Y., ranging in education from seven years of schooling to graduation from college. Many nationalities and some Negroes were included. The testing administration and scoring were uniform and were done by the same person. The Hawk-Eye 16-25-year age group scored higher on all but one percentile level than the published norms. The Hawk-Eye 26-60 group was never higher than the 16-25, though on the published norms this condition existed at the 5th, 10th, and 15th percentiles. The authors believe their data may be more representative of adult female workers than the original norm group. Elizabeth Bell

Bingham, W. E., Jr. "A Study of the Effect of the Presence of the Examiner upon Test Scores in Industrial Testing." *Journal of Applied Psychology*, XXVII (1944), 471-477

Thirty-six men and 24 women averaging 21 years of age, all college students, were given equated forms of steadiness, typing, and addition tests in the examiner's presence and alone, the purpose being to check the validity of mechanical ability and personnel procedures where the job is a comparatively solitary one. Results

* Edited by Forrest A. Kingsbury.

indicated that the subjects were more efficient in each test when working alone, although in the addition test they completed more work with the examiner present. Introspection showed the examiner to be the stimulus factor for these effects. Further research is needed, according to the author *Vernon S Tracht*

Blain, I. J. "The Rationale of Scientific Selection" *Occupational Psychology*, XIX (1945), 28-34.

This article supplements that of Cockett's, re-emphasizing some of its points and introducing others that had not been previously mentioned, all in line with the application of scientific techniques to the problems of personnel and job-placement in industry. Psychological tests are seen to give the employer a more objective, critical means of appraising a person's abilities than application forms, school reports, and interviews, although the latter are recognized as contributing information concerning personality, character, and temperament not as yet accurately measurable *Vernon S Tracht*

Burton, Arthur and Joel, Walther. "Adult Norms for the Watson-Glaser Tests of Critical Thinking" *Journal of Psychology*, XIX (1945), 43-49.

The *Watson-Glaser Tests*, Battery I, were administered to 150 applicants for civil service positions. Ages ranged from 23 to 72 years, with the mean 39.1 and median 36.9. Subjects below the median age made significantly higher mean scores on all tests. The mean score of those with 2 or more college degrees was higher than those with 1 or no degree, but the difference was not statistically significant. Norms as a whole were higher than norms for college seniors. More extensive norms are needed, and a study of the validity of the tests in selection of professional and administrative personnel should be made. *Lorraine Bouthilet*

Cockett, R. "The Rationale of Scientific Selection" *Occupational Psychology*, XIX (1945), 20-27.

Contrasting the autocratic with the democratic way by which a society may fully utilize the abilities of its individual members, the author describes the process of "scientific selection." This involves the organized method of discovering which persons possess the skills and capacities necessary for certain kinds of work, and of analyzing different jobs to determine what abilities are required for them. The experimental and statistical methods of securing valid and reliable measuring instruments for this purpose are briefly discussed and the personal or "human element" assessed *Vernon S Tracht*

Goldstein, H. "A Malingering Key for Mental Tests" *Psychological Bulletin*, LXII, 104-118.

The malingering key is a scale composed of those items which proved most sensitive in differentiating between simulated malingering and genuine failure on the Army's *Visual Classification Test*. It is applied directly to the original test papers and yields a score based upon the number of discriminating easy items failed and the difficult items passed. The key is developed upon the hypothesis that morons and malingerers give test patterns differentiable because bona fide failures will fail the harder items, while the malingerers will tend to pass more of the hard items than the bona fide failures. In trials with thousands of cases, the key eliminated from seventy-five to ninety per cent of test failures as non-malingerers, thus enabling examiners to concentrate upon the remaining few and prove whether they were mentally adequate for service *Elizabeth Bell*

Jurgensen, C. E. "Report on the 'Classification Inventory,' a Personality Test for Industrial Use." *Journal of Applied Psychology*, XXVII (1944), 445-460.

Designed to avoid some of the faults of personality tests now used for predicting job success in industry, this inventory contains 245 items, comprising 45 groups of three items each and 55 paired comparison forms, and is intended to be scored and validated on jobs instead of personality traits. The author asserts that keys may be devised on the basis of traits for utilizing it in industrial, educational and clinical guidance work, but that occupational selection requires keys based on

specific jobs. Reliability and validity have been "satisfactory" in the two studies thus far completed, validation having been on groups comparable to those for whom the *Classification Inventory* is planned. *Vernon S. Tracht*

Lowenfeld, V. "Tests for Visual and Haptical Aptitudes." *American Journal of Psychology*, LVIII (1945), 100-111

Having determined that there are two groups of individuals with respect to their orientation a) those who use their eyes as the main intermediaries for their sense impressions and b) those who, though with normal sight, depend upon touch and kinesthesia, the author has developed a battery of tests for use in job selection, where either the presence or absence of these abilities may be significant. On an experimental group of 1128 reactions, it was found that 47% were clearly visual, 23% were clearly haptical, and 30% were not distinguishable. Stress on this knowledge is important in job situations where the presence of one or the other aptitude would be a liability. *Francis Medland*

McHugh, G. "Relationship between the Goodenough Drawing a Man Test and the 1937 Revision of the Stanford-Binet Test" *Journal of Educational Psychology*, XXXVI (1945), 119-124

The author found an r of .45 between M.A. scores and an r of .41 between the IQ scores on the *Goodenough Drawing a Man* test and those on the 1937 *Stanford Revision* (Forms L and M) of ninety kindergarten children in the public schools. These r 's, he believes, may be depressed because of the equal use of both forms of the *Binet*. A further study of the bi-serial r 's between individual scores of the two tests reveals that limiting the *Goodenough* to the nine items which have a biserial correlation coefficient of .30 or better with *Binet* IQ produces the best relationship between the two tests. Thus the number of *Goodenough* items used for kindergarten could be much reduced without losing reliability. *Elizabeth Bell*

McNamara, W. J. and Weitzman, E. "The Effect of Choice Placement on the Difficulty of Multiple-Choice Questions" *Journal of Educational Psychology*, XXXVI (1945), 103-113

It is generally believed that the "chance" element in "five-choice" and "four-choice" objective questions is "one-in-five" and "one-in-four," respectively. This study attempts to determine the possibility that placement of the correct choice in one of the four or five possible positions has a definite and measurable effect upon item difficulty. On a group of Naval Cadet subjects the next-to-the-last position, on both four- and five-choice items, had the greatest difficulty level. On four-choice items, the difficulty increased from the first through the third position. In five-choice items the second and third positions were less difficult than the first, and the fifth position not significantly more difficult than the first. The data also indicate that the act of reading several incorrect choices has little or no effect upon the ability of the subject to select the correct choice in the series. *Francis Medland*

Schmidt, H. O. and Billingslea, F. Y. "Test Profiles as a Diagnostic Aid: The Bernreuter Inventory" *Journal of Abnormal and Social Psychology*, XL (1945), 70-76

These Army psychologists found that when subtests B1-N, B2-S and B4-D of the *Bernreuter Inventory* were considered in relationship to each other by statistical treatment, they were able to differentiate "standard normals" from "standard deviates" with an approximately 80% degree of certainty. Enlisted men of the Army Air Forces made up the 2 groups of subjects, 100 in the normal and 329 in the maladjusted, the latter having had psychiatric diagnosis by competent medical officers. While not indicating the direction of maladjustment, the inventory is believed by the authors to be a time-saving method of spotting deviate individuals. *Vernon S. Tracht*

Slater, Patrick. "Scores of Different Types of Neurotics on Tests of Intelligence." *British Journal of Psychology*, XXXV (1945), 40-42.

This paper defends the "theory of overlapping group factors facilitating neurosis." Twenty-five men of each type of neurosis, obsessional, miscellaneous, anxiety

states and hysterias, were tested over a two-year period. The main tests were *Progressive Matrices*, *Castell IIA* and *IIB*, and the *Shipley Vocabulary Test*. Those with obsessional neuroses tended to be more intelligent than those with the other types. There was no proof that the others differed significantly, in other words, "neurotics are heterogeneous as regards intelligence." It is argued that if we find many common characteristics which differentiate neurotics from normal persons, we can expect also to find, when we isolate a particular characteristic, that the "neurotics are heterogeneous in its respect." *Elizabeth Bell*

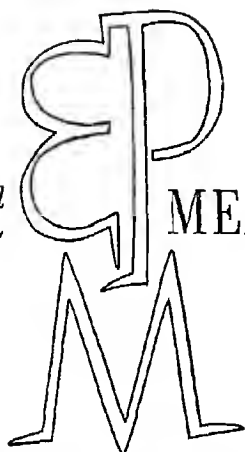
Staff, Psychological Section, Office of the Surgeon, Headquarters, AAF Training Command, Army Air Forces. "Psychological Activities in the Training Command of the Army Air Forces" *Psychological Bulletin*, XLII (1945), 37-54.

This is the seventh of a series of articles and is a report of the activities of the Section in the "application and correlation of the various tests used in the classification of aircrew members" and in the supervision and coordination of psychological research activities, including test development. Emphasis has been shifted from selection of aircrew trainees for success in training, which has been practically solved, to the selection of those who will make good combat officers. Data are being secured against which the tests may be validated. Efforts are being made to select men for the more specialized functions as members of lead crews, fighter pilots, and bomber pilots, and for the various types of gunnery training. There is also an extended search for proficiency criteria in training. *Elizabeth Bell*.

Wilson, G. M. and Burgess, Faye. "Construction Puzzle B as an Ability Test" *Journal of Educational Psychology*, XXXVI (1945), 53-60.

This is a discussion of the results of using *Construction Puzzle B*, a form-board test, as part of a battery of tests given in a war industry for classification purposes. It was observed that some subjects who were low on the form-board test were high on other tests. The need for special attention and interpretation, as well as the need for further study, in regard to this test is emphasized. *Lorraine Bouthlet*

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME FIVE, NUMBER THREE, AUTUMN

<i>Some Concepts of Job Families and Their Importance in Placement.</i> HERBERT A TOOPS	195
<i>Testing for Administrative and Supervisory Positions.</i> MILTON MANDELL	217
<i>Army General Classification Test Scores for Civilian Occupations.</i> THOMAS W. HARRELL and MARGARET S HARRELL	229
<i>The Nature of Mechanical Ability.</i> J. R WITTENBORN	241
<i>Suggestions for the Construction of Multiple-Choice Test Items.</i> CHARLES I. MOSIER, M. CLAIRE MYERS, and HELEN G. PRICE	261
<i>Influence of Level of Motivation on the Validity of Intelligence Tests.</i> L. D. HARTSON	273
<i>Statistical Interpretation of Symptoms Illustrated with a Factor Analysis of Problem Check List Items.</i> STANLEY S MARZOLF and ARTHUR H. LARSEN	285
<i>New Tests of 1944-1945</i>	295
<i>The Contributors</i>	301
<i>Membership List</i>	305
<i>Measurement Abstracts</i>	315

The Contents of This Issue Are Listed in the *Education Index*

PRINTED IN THE UNITED STATES OF AMERICA
THE SCIENCE PRESS PRINTING COMPANY
LANCASTER, PENNSYLVANIA

SOME CONCEPTS OF JOB FAMILIES AND THEIR IMPORTANCE IN PLACEMENT

HERBERT A. TOOPS¹

Ohio State University

THERE are, it is asserted, some 30,000 or more occupations. That is a large number. We could hardly construct that many aptitude tests, say, in a century. To any other proposal affecting all the occupations one would have to make a similar comment. The task is too big; it would not get done. There is great need, accordingly, for somehow reducing the number of "kinds" of occupations. Could one sort out, for example, a small number of type-occupations which would stand for or represent the lot of them?

This hope is analogous with the corresponding dream of psychologists regarding human types. They hope to be able to type all humanity into a relatively few "unique personality profiles," or patterns. Thus, though the people in a given type still would differ considerably amongst themselves, such differences might be thought of as relatively unimportant. The people of a given type, however, by definition would be singularly alike in, say, such matters as aptitudes, health, drives, wants, and satisfactions. They might still differ greatly in race, color, religion, height, weight, appearance and other respects.

The importance of such maneuvering is easily seen. For example, in any psychological, sociological or educational study, "like persons" must be subjected to the test of the experiment. With only a limited number of types of people to be concerned about it would be easy then to get together

¹On leave with the National Scientific Roster of Scientific and Specialized Personnel. The writer is indebted to Dr. Beatrice J. Dvorak, who made consultation with WMC officials possible, and to Dr. Carrol J. Shartle for helpful criticism of the article in formulation.

enough persons for a study; enough that any statistical follow-up would be statistically reliable. Without some such conception as, "unique personality profiles" every person obviously has a profile so "complex" that he is his own type, thus resulting in 140 million such types in America, and no studies are possible.

Just as the botanists find it useful to group their plants for study, and the zoologists their animals, so must psychologists find a way to group their humans. How to do this is the question. This they may not do by employing the United States Census classifications. The psychological differences, for example, are small between whites and Negroes; between men and women; or the single and the married. We must search deeper for the more fundamental differences among individuals, indeed for the traits which are the "basic dimensions" of humans.

The traits on which unique personality profiles are founded must be chosen—it is known—to obey such general principles as:

1. The traits must measure basic human "dimensions" or qualities. Another way of saying the same thing is to state that:
 - a. The traits, in the aggregate, must correlate highly with success in all the important endeavors of mankind, particularly the occupational success criteria.
 - b. They must correlate zero, approximately, with one another to the end that each such trait thus measures as much as may be something different from every other, a basic human "dimension."
2. The number of such traits to be measured will therefore become a minimum; and the "minimum profile of personality" results.
3. The traits must be objective and quantitative, and (preferably) not too expensive of time or resources for their accurate measurement; in a word, they must be capable of being measured by practical tests or measuring instruments.

Obviously analogous concepts underlie any attempt to classify occupations. The "traits" of occupations likewise must pass the same standards. They too must be unique, minimal in number, and as objective, quantitative and practical as possible. There is here, however, an additional requirement, popularly referred to as a means of "bridging the gap between the man and the job." Briefly this means that, if possible, the occupation should be measured in the same units as the man, to the end that we may form reasonable judgments of whether this given man can do this particular job; or rather, since the man is the reference point rather than the job—or popularly, since industry exists for man and not man for industry—which job, of the myriads of jobs, can a man of this particular profile in general do best? The reference point often is reversed, and particularly in wartime where filling the "openings" is the paramount consideration, even though it does not follow in every case that the individual "works at his highest skill." War is an emergency; it is met with emergency behavior. Thus in wartime we may find it desirable and necessary to employ practically all 24-percentile, or higher, chemists as *chemists*; but we also may employ even 100-percentile lawyers as clerks or infantrymen. (The professed aim of course is not to do this, but, after all, any soldier is *one* soldier!) With peace, the guidance consideration will again assume its former relative importance.

With humans and jobs measured in comparable units, and having amassed in addition a very abundant evidence upon the efficiency of each given type of man in each given type of job, very vigorous modes of sorting men to fill jobs (for selection or placement) and of sorting jobs to suit men (for guidance) at once become possible. The technical problems therein are at least partly solved. With the two measurement problems solved, the problem of measuring the correspondence would readily yield to attack.

Hitherto this correspondence of job and man has been "decided" by a counsellor, employment clerk or placement officer, customarily all too ignorant of the varieties of human nature, on the one hand, and of the varieties of jobs on the other. And

customarily he has thought of the job as a different-order-of-creation, with its own traits and categories that bear little or no relation to human traits and categories. Clearly then any mode of classifying occupations and jobs which will emphasize such values and relationships—particularly any which will make the relationship obvious—will be valuable and likely will lead both to better selection (placement) and to better guidance. Ideally, if there could be a one-to-one correspondence between the occupational “traits” and the man traits, placement would be maximally facilitated. Possible methods of making this correspondence are considered herein. Briefly, they are:

1. Stating the requirements of the occupation in terms of the average or, preferably perhaps, the 75-percentile as-to-success worker therein, on the traits on which it is useful—i.e., valid—to measure the workers.
2. Specifying an occupation as an idealized set of human requirements, or profile of human traits, based on human judgments of the occupation’s “requirements.”
3. Specifying all occupations in terms of the human traits which a factor analysis of data, obtained preferably by method 1 above, would reveal as the common measures of all jobs.
4. From empirical observations of success on jobs of a very large number of men of varying profiles, specify as the human-characteristics profile of the job, that human profile which succeeds best at the occupation in question.

The attempt to classify occupations has so far led only to the concept of job families. The concept has grown out of employment office work, where in normal times they have on hand unemployed men needing work. Their search at such times is for an available job which a given man, say an accountant, can do. Their special knowledge consists in convictions of what a given type of man can do.

Our problem is to increase that knowledge and to amplify its dependability. During the depression it became obvious that an accountant could do labor work, if he had to, in order to live. Even depressions, however, do not greatly increase our knowledge of what *more complex* work accountants can do.

Clearly, if we could systematically accumulate such experience for all or most men of the community as they change jobs, we should soon have a vast fund of knowledge for sifting and refinement. In wartime many drafted men arbitrarily and by chance are placed in occupations new to them, but in most cases these too are down-grading; and furthermore, little record is made of the satisfactoriness of such placements. At such times much job-dilution (job-simplification) is resorted to in an effort to employ less capable talent on small units or specialized aspects of more complicated work, to improve production and to increase accuracy by the development of great individual skill on non-complex tasks.

It has long been known in employment office work that when industry could not get "exactly the type of man it wanted," a man of "certain specific occupations" often might do almost as well. In Swan's Index, for example, prepared for the C.C.P. personnel work of World War I, the "substitute occupations, or civilian equivalents," of some 750 Army trades and occupations were determined and listed for the use of all personnel officers. The problem, it has been stated, is that of how to transfer workers from job to job with the fullest possible utilization of their previous skills. Thus the Army should exploit the civilian occupational skills of the draftee, and in turn industry, after peace, should exploit the Army-acquired occupational skills of the demobilizee. Inasmuch as any definition is always defective, let us enumerate the possible characteristics of a "job family," by implication, from a consideration of the actual or proposed practical uses of such "job families":

1. When up-grading, training will be more effective if it is for a new responsibility which falls in the same family of occupations. Thus shortages of manpower may be readily filled from excesses of manpower if such exist anywhere in the same family of occupations. By such action the amount of re-training will be minimized. (It will be recalled that different occupations turn out different products and these are peculiarly affected, in wartime particularly, by the presence or absence of raw materials. Accordingly, "excess" of manpower may

crop up almost anywhere after every new governmental order or decree.)

2. For determining substitute occupations for individual employment purposes, where the additional training, if any, is to be afforded by the job itself, job families provide the necessary factual information. The greater job families, based on industries rather than occupations, enable wise decisions to be made in converting non-essential industries to essential industries, with a minimum of waste.² The complete present manning tables of a factory compared with the proposed manning tables of a new production schedule reveal at once the distribution of occupations freed and those necessary for operations after the conversion. Let the distribution of occupations released be abscissae (X) of a two-way table; let the proposed new distribution be ordinates (Y) of the same table. The conversion is effected then in the following series of steps:

2(1). Into the diagonals of the table, which have identical labels for each coordinate, are placed the several frequencies of persons who can be transferred directly without conversion, into exactly the same occupations. The personnel cards, or duplicates thereof, of these specific persons are then sorted from the total work-roster and are filed under the new manning categories.

2(2). In each diagonal compartment, form a ratio of the number thus obtained to the total number desired in each of the several occupations. Parallel this record with another of arbitrary indices of indispensability of the unfilled portions. By a study of these two ratios, giving weight to the importance of the new occupations in the scheme of production, and to salaries, training time, and possibility of obtaining outside trained recruits, establish an order for filling the new occupations. It would be highly desirable if this angle of the matter could be made objective and mathematically determinate.

2(3). Reduce the X-marginal frequency, or abscissal marginal row of the chart, by the frequencies of persons just allocated (Step 2(1) without conversion).

² Shartle, C L, Dvorak, Beatrice J. and Associates "Occupational Analysis Activities in the War Manpower Commission" *Psychological Bulletin*, Vol XL (1943), 703 Shartle, C L et al "Ten Years of Occupational Research" *Vocational Guidance Magazine*, XXII (1944), 387-448.

2(4). Considering now the new occupation (Y) to be first brought up to strength: by aid of the job family classification pick out the *first* conversions therein, if any, those which in general can be made with a minimum of retraining. Add the frequencies of the corresponding persons to the compartment which is designated by an X of the to-be-invaded occupation and a Y of the to-be-established occupation, somewhere beneath the diagonal row of compartments; and remove a corresponding number from the X-marginal frequencies and also from the remaining roster the cards of those so related and converted. Similarly invade the other earliest-to-be-invaded occupations; and when all is complete establish in a second Y-marginal column the frequency of the most indispensable of the new occupations. If the quota is filled, the process is stopped short at that point, only enough persons being taken from the last occupation to fill the need.

2(5) In similar fashion, consider the second occupation to be brought up to strength, the third, and so on until all the possibilities of *first* conversions are exhausted, or until the quota is filled, whereupon the sorting of cards is at once stopped. But if this does not yield enough personnel, resort next to the *second* conversions, those requiring more training, and to the third and higher conversions, if any, until either the quota is filled or there are no more possibilities of conversion remaining. The several conversions may be kept separate by different colors of ink, and are a training order in detail to the training department: "You are to train so many lathehands to be assemblers, etc."

2(6). Consider next the second most urgent new occupation, repeating steps 2(4) and 2(5).

2(7). So continue until there are no more conversions to be considered.

There will now be two groups of interest:

A. A residue of personnel not useful to convert. If there are shortages in other areas of production, these may be subjected again to a second scanning with the end of filling these shortages.

B. Shortages in the proposed new manning tables which must be filled by resorting to outside recruitment.

Clearly there is room here for the development of mathematical methods of optimizing placements.

3. Transferring an unsuccessful person within a firm into an occupation in a job family in which he had formerly been successful, as an alternative to separation from the firm, is wise personnel management.

4. In giving guidance to individuals who wish an improvement in their situation but are unable to forego an income while retraining, job families are helpful.

It is a well-known principle of adult guidance, for example, that if possible the guidee should exploit his most successful experiences and training rather than merely abandon them when he changes occupation.

5. In recruitment of personnel for new industries, job families inventories yield valuable information as to *where* such recruitment will yield the best results at the least cost.

6. In the assignment of recruits to related military occupations, job families prepared for the Army, Navy, Marines and U. S. Coast Guard have been helpful.⁸

7. That wages of the member occupations of job families, other things being equal, should differ but little, is a principle of value in wage adjustment.

8. The various occupations of job families probably have highly similar physical demands, thus making it possible to know alternative positions into which a given handicapped person will fit, without preliminary research or trial and error.

9. In promotions one wants to maintain the principle of individual growth, rather than that of revolution, so that a knowledge of job families should be helpful in establishing paths of promotion. One is ready for promotion when one is proficient in his present job in all the elements thereof that are common to the next job ahead, this assuming that the paths of promotion previously have been established according to such principles. One's present job is thus always a production field for exercise of the skills, knowledges and tech-

⁸ Shartle, C. L. *et al*, *op cit*, p 704

niques obtained in the previous position, and a training field for acquiring those skills, knowledges and techniques basic to the production of the job ahead which also are a part of the reasonable practice of the present position.

10. In aptitude test construction, a test conceivably may be devised which will be reasonably adequate for all the occupations of a job family, thus restricting and constricting the field a great deal.

11. The reabsorption of veterans and the shift of civilians from wartime production to peacetime production after peace and the demobilization will involve a stupendous task of worker transfer, retaining and allocation. As an aid in this task, job families, applied in reverse, will be useful to locate the logical job destination of workers in civilian industry when war and war industry no longer exist. In theory this is analogous with the construction of a decoding code. Inasmuch as it is reasonable to believe that many demobilized at peace will be ready for promotions, their after-war destiny logically also should be a related job-family occupation to which a given worker "should be promoted" in view of his increased "skill" acquired during the war years. The longer the war, the less advisable it is for a larger and larger number of the demobilizees to return to their former jobs

The common statistical element in all the above uses of job families is that a few "type" jobs may stand for all jobs just as a few personality profiles may stand for all human profiles. The end to be secured is a great reduction in the magnitude of the problem and in the amount of data or evidence which must be obtained to solve a problem.

The War Manpower Commission says that there are six classes of requirements⁴ on which occupations should be compared, namely,

- (1) Nature of the work done.
- (2) Tools, machines and other aids employed.
- (3) Materials worked upon.
- (4) Traits required of the worker.
- (5) Knowledge (including specialized knowledge) required.
- (6) Experience.

⁴ Stead, W H and Shartle, C L., *op. cit*

Occupations grouped on the basis of the work performed (i.e., sorted on the basis of action verbs, such as welding, sewing, nailing, gluing, etc.) seem more alike than if grouped on any other basis. In the United States Employment Service sorting of the "traits of occupations" to determine the job-families, the first four of the above list occur in the order above named in the sorting operation. Shartle's results seem to show that "people," rather than specific kinds of people, are required for a majority of the occupations of industry, and that the training time is, for the great majority, unbelievably short. Wartime training methods of unusual potency emphasize the correctness of this conclusion. Motivation was often at a maximum. If you did not learn, you got your head shot off!

As a means of facilitating their use, the coding scheme employed to stand for an occupation should take into account the family relationship of occupations by ascribing contiguous numbers to the highly related occupations. Occupations thus break down into "fields of work," somewhat narrower "process groups" and, finally, into still further variations, varieties, alphabetically arranged (and distinguished by unit digits in the USES code number).

Just as in the taxonomy of botany one may arrive at various conceptions of what is a "family" (a sub-division of higher divisions of classification) so one may arrive, by different routes, at different classification principles, at different aggregations of occupations which in the several classification systems logically may be called families. Some half dozen such alternative, actual or potential systems of deriving families of occupations will now be outlined.

1. The most extensive development of job families has been made by the technique developed by the WMC, Worker Analysis Section. It involved the following steps:

- 1(1). A job characteristics form, consisting in its final edition of 47 human traits, was filled out by from one to fifteen analysts in different parts of the country observing the same occupation, as distinguished from job-in-this-particular-factory.

- 1(2). The same analysts also observed the duties of the

worker, the tools used, the equipment, the materials and the minimum education and experience necessary for the work. The results were systematically recorded in a Job Analysis schedule form.

1(3). A compromise set of "traits" for the given occupation was decided upon, and was entered into a Master Worker Characteristics Sheet.

1(4). The results now were recorded into Speedsort cards and other pertinent observations about the occupation were written on the face of the card.

The cards now represented in highly condensed form the results of judgments of what are the important characteristics of the occupation in question. Some 9,000 occupations, in some 85 industries, were thus reduced to Speedsort cards, a kind of catalog of occupations which could ultimately, after the job is finished, become *the* card catalog of American occupations. By the aid of the usual sorting needles, presumable families, having regard for the above list⁵ of judged most-important "traits," were sorted out. This was aided by noting that as the sorting progressed, channels in the edges of the cards began to develop on other traits of a family than in those traits actually sorted upon, the traits of the list above. Aided by these, the member-occupations of presumable job families were sorted out.

The complete job analysis description was appealed to as a further test for excluding some of the presumable members of the family; and to insure that all that should belong to a family had been included. The job family was now complete.

1(5). There was still a need for ascertaining which of the members of the job family were most like the ideal "job type" from which all the occupations of a family deviate in some respects, even "important" ones. By a highly subjective procedure, based first on a guess as to the relative importance (weights adding up to 100) of the half-dozen or more generalized aspects⁵ of the type occupation, and second, upon the subjective points-rating therein of a given occupation, never

⁵ The aspects as above outlined are: (1) work done; (2) tools, machines and other work aids used; (3) materials used, (4) worker characteristics required, (5) experience required; (6) training (including *special training*) required

exceeding the previous "weight" of this particular member-occupation of the job family, there was ascertained an aggregate numerical value (the sum of the points-ratings) by means of which the specific occupations could be ranked in a decreasing order of degree of resemblance to the type job.

By aid of these, the job-family was split into three subdivisions:

Closely related occupations—those in which workers of experience could probably be transferred to any other member-occupation of the closely related group with little or no retraining.

Less closely related occupations—those for which workers required more retraining to make them acceptable workers.

Least closely related occupations—those barely meeting the minimum requirements for useful similarity, and for which workers, although they might require complete retraining, nevertheless because of their worker characteristics and other considerations of knowledge, skills, etc., would be more likely to succeed at the work than just anyone chosen at random.

The apparent strong points of the method are:

- a. A wide variety of "traits" of occupations was investigated.
- b. The analyses of the job were made by trained analysts.
- c. Observations of the same occupation were made in different parts of the country so that sectional discrepancies could be allowed for, even to the extent of splitting up an ostensible occupation into two or more; or it could be noted that occupations bearing differing names in different parts of the country are in reality one occupation, not several.
- d. A common denominator of the several alternative analyses was decided upon
- e. The job families resulting had other than statistical and logical definition and justification, since the sorting of the Speedsort cards was augmented by a

critical scanning of the entire description of the constituent occupations before adoption.

Its more obvious weaknesses are:

- a. The analyses were subjective, even though done after a uniform schedule or outline.
- b. The list of job "traits" was not wide enough to cover adequately the professions.
- c. The comparison of the reports of the as many as fifteen analyses of a given occupation was done centrally by a person out of touch with the field and without giving the analysts who did the work a chance to correct the most obvious misconceptions. The completed analysis was not generally sent to them for criticism before official adoption, although some one or more analysts scanned each of the proposed analyses before adoption.
- d. Different Speedsort operators, with the same end in view, presumably could come up with different potential job family-occupations. In other words, the system was not fully statistically determinate.
- e. There is some doubt whether the skills and traits of workers are in every instance transferable merely because the pattern in the Speedsort cards is highly similar. A watch-maker and a cannon-barrel borer might come out in the same job family and on the basis of the statistics have every warrant for belonging to the same family, yet the psychological characteristics, particularly as to precision, may be quite different so that actually there is little transferability of skills.
- f. The double dose of subjectivity involved in putting the sub-members of a job family in decreasing order of resemblance to the common type is unfortunate when there exist several alternative techniques, independent of subjectivity, which will give a numerical measure of the degree of correspondence of two profiles. The ideal jobs or type jobs we take to be a pure figment of the imagination. It has no

more objectivity than anyone's conception of the "ideal mammal," the "ideal liliacea" or any other taxonomic classification. In fact it does not have the rigidity of definition of a sub-classification of botany in that *all* liliacea are spermatophyta (seed-bearing, with true roots, stems and leaves), angiosperms (with true flowers, each containing, normally, four whorls of floral organs), autophytic (possessed of chlorophyll), monocotyledonous (one cotyledon in each seed and with parallel-veined leaves) and bulbous (having bulbs).

2. Inverse factor analysis may be employed as an alternative means of ascertaining a job family. If the characteristics of an occupation can be quantified and measured, a factor analysis—occupations replacing the usual human names—will reveal what factors of job-traits underlie occupations in general. In this there is no necessary restriction of "the occupation" to any particular pattern of measurements, for these may be quite as broad as, or broader, than the WMC list above, and in the case of the professions⁶ almost surely will be broader than that list! Occupations having similar factor loadings belong to the same family.

3. A variant on method 2 above emphasizes successful plying of an occupation and implicitly recognizes that every occupation has at least a fringe of people "who ought not to be in the occupation." If the various X's of the above method are replaced with a sufficiently varied list of the human traits of, say, "*successful workers*" of the occupations in question, the resulting factors will discover the basic human factors of workers. In this case also one may express the occupation in terms of its component factors and factor loadings. The loadings, here, are the profile of the occupation; and those occupations with highly similar profiles are jobs of highly similar human requirements. The occupations which resemble each other most comprise the job families.

⁶ Professions are generally characterized by a higher level of job-traits than characterizes occupations which are not professions, in particular in such respects as research, administration and supervision, and knowledge. There is probably no job-trait found uniquely among professions save that of being dubbed a "professional."

The occupations of a given job family presumably may be located after the occupational profile is obtained by such methods as 1 or 2, either by (1) the Sagebeer Index, or alternatively by (2) the Stephenson inverse factor method. In employing the Sagebeer Index, a rationale of research for locating families would need to be developed.

4. A fourth method might be dubbed the universal follow-up method. If one kept track of all or nearly all placements of professional personnel and at the conclusion (resignation, discharge, transfer) of the jobs of such personnel collected a simple verdict as to whether or not the person in question had been "satisfactory" in the position in question, one might analyze the resulting data to note by a simple correlation index which occupational transfers are successful and which are not. With all the occupations of concern appearing in alphabetical order as both the ordinates and the abscissae of a two-way table, one may let Y be prior job and X be subsequent job. In the compartments thereof, let four-fold correlation coefficients, solved by the formula,

$$r = \frac{ad - bc}{\sqrt{(a+c)(b+d)(c+d)(a+b)}}$$

be recorded. (See Fig. 1 below.)

		Subsequent Job (X)		
		Unsuccessful on subsequent job	Successful on subsequent job	Total
Prior Job (Y)	Successful on prior job	c	d	c + d
	Unsuccessful on prior job	a	b	a + b
Total		a + c	b + d	N

FIGURE 1

Four-fold Validity Plot of Success on Prior Job and on Subsequent Job
(If success can be more adequately measured than implied in the bifurcation illustrated, this four-fold plot would be replaced by an $m \times m$ -fold plot)

The letters of the diagram (Fig. 1) are the frequencies of the table, here recorded literally to show the mode of computation by the formula. From such a table of τ -indices, by means of a suitable index to be devised, one could readily sort out the job families. It will be noted that one obtains here two-directional indices. Thus, not only would one consider the question "Are the failures of job A also failures when they subsequently enter B?" but also "Are the failures of B failures when they subsequently enter A?" If the two tendencies are vastly different, as in some cases they may be, we might discover here some preferred *sequences* of progression in acquiring vocational skills.

It is normal for wide-awake, alert human beings to take on more and more skills, including jobs and even occupations, as they develop and mature. We know little enough about the more complex end patterns characteristic of "the expert." One interesting question here is whether the expert usually is characterized by having acquired his complement of skills in an orderly progression. If the answer were in the affirmative, obviously one might similarly order the acquirement of skills of all people to very good advantage. This method logically requires that all the experience of a large community of occupations be amalgamated, say of a WMC area at least.

5. The logical or evolutionary concept of job families, described by Knott,⁷ conceives that jobs which belong to a family are the variations which a specialized industry has evolved out of a simpler craft. Thus hat-making, cap-making, dressmaking and even baggage-making in common employ variants of the power sewing machine (both the tool and the corresponding skills), itself an invention superseding in turn hand-sewing. And typists, stenographers and secretaries are variants of the amanuensis or letter-writer who still plies his skills in countries where most of the inhabitants are illiterate. In such variants some retraining is needed before an individual successful in one mode of sewing may become proficient in another, but, obviously, *the retraining period would likely be much less* than if

⁷ Knott, Edward E. "Job and Occupational Analyses" *Missouri State Employment Service*, April, 1941, p. 47

a highly skilled operator in any one of such were to be trained into a highly skilled operator of, in general, any other job not included in the job family, say a lathe operator. Thus an index of what the educational psychologists call "transfer value" would seem to be at the heart of this concept, which is based on logical, that is economic, historic, or evolutionary, concepts of the mode of development of the occupation in question. Insofar as such families can be ascertained without the necessity of the empirical evidence of the #4 method above—empirical ascertainment of transferability⁸—the method serves usefully to restrict the field of inquiry but does not eliminate the need for the previously mentioned modes of study. There are sub-families within families and sub-sub-families within sub-families. These may be ascertained more surely by statistical analysis than by unaided observation. One strength of the method is that more than the one source of data, the customary field observation of jobs by trained workers, is insisted upon as necessary to a proper classification. (To properly classify plants, one must study paleobotany as well as botany.) With a set-up such as No. 1 or No. 2 above, the sub-occupations, members of eventual potential sub-family groupings, would be the Y's of a basic data table whereas the X's might be either (a) tools employed, (b) unit-operations employed, (c) motions, in the Gilbreth sense, employed, or (d) worker characteristics, or any compound of the several kinds of such components, or (e) any other of the WMC classes of observations. The subsequent analysis of the collected data would follow that of Nos. 1 and 2 above.

6. An alternative to all the preceding is that developed by Mr. Robert Shosteck and associates of the Roster of Scientific and Specialized Personnel, to meet the emergency in prospect in the sciences when demobilization occurs. It is simply that of writing to industrial experts employed in practical work, including the selection of scientific and specialized personnel, and

⁸ Evidently the indices of success of the basic table of the No. 4 method might be replaced by indices of amount of trainability required, say the average of retraining actually administered, it being assumed that retraining is stopped as soon as proficiency is achieved.

asking them to give their judgment as to the degree of transferability—on a scale:

1. *Considerable* transferability (Less than six months additional training required)
2. *Moderate* transferability (Six to twelve months additional training required)
3. *Slight* transferability (Over twelve months additional training required)
4. *No* transferability (Transferability unfeasible, since little or no training time would be saved)

—of people in a given series of occupational specialties of, say, civil engineer, when and if these were to enter the break-down specialties of the same occupation or the specialties of another industry, say the engineering specialties of municipalities. The present hypothetical occupations of the to-be-released workers are Y-ordinates of a two-way table, while the X-abcissae are the jobs of an engineering sort now maintained in municipalities. In this case, hydraulic engineers, for example, were asked to judge on the above scale the transferability of each of a number of kinds of civil engineers to the several other specialties of the same occupation. The validity of the method hinges on the ability-as-judges of the addressees, i.e., upon such factors as the extensiveness of their actual knowledge and experience with the implied situations judged, the criticalness of their judgment, and their willingness to take the necessary time—all of which could no doubt be brought to a state of excellence by well-known techniques, such as that of using lesser experts to pick the greater, and the like. The method has the merit of quickness, low cost, and independent verification of the observations exploited. It has all the weaknesses of using untrained judges in what is essentially a psychological experiment. It may be pointed out that the WMC job analysts also employed judgment, of small elements of an occupation to establish a profile from which, by statistical manipulation, the job family ultimately was determined and the transferability of skills was inferred. Here, however, the transferability of the specialty as a whole is judged. By abstract psychological principles the judgment required here is easier to obtain but

is of lesser validity than judgment of the more specialized sort obtained in the former studies. For the aid of demobilization counsellors, the less costly method yields results which presumably are about as fine as demanded for the purpose. The finer break-down of the WMC job analysis obviously enables more uses to be made of the results, such as advising the demobilizee to know what specific education or training he should acquire for a better all-round fitness in his new vocation. This the Roster technique, with its present paucity of data on the individual profession, cannot supply. However, having helped the demobilizee to choose wisely a new occupation which will exploit his army and pre-army training, such additional (educational) information in many cases may be had even more usefully from a new counsellor, perhaps a dean or secretary of an engineering college. The advisee is less condemned by receiving only parts of the truth than by receiving mistruths. If the validity of the method is equal to or only slightly inferior to the more laborious method, obviously then, on this score alone, it has great merit. Incidentally the statistics of ratings, for subsequent treatment of the results, are already well worked out. The necessary statistical indices in no case are an insuperable element of the whole. The rating scheme for rating officers' traits in World War I was deemed a rating failure by its chief statistician.⁹

Confronted with such a variety of actual or potential techniques, one feels a need to inquire, "Which is best?" The question implies the existence of some standard, some criterion, by means of which one might adjudge relative merit. Insofar as any such system is, from one angle, merely a filing or classifying system, obviously all the systems solve at least minimally the problem in the sense that every occupation has a position, and a code number, and none is omitted. However, a classification system may do and usually does do more: it leaves gaps to be filled in by variant and newly emergent classes. For some purposes the gaps may be quite as important as the filled niches.

⁹ Rugg, H. O. "Is the Rating of Human Character Practicable?" *Journal of Educational Psychology*, Vol. XII (1921), 425-438, 485-501; *et al.*

The common purpose of job families, if the multitude of purposes enumerated above can be subsumed under one general statement, is that of ascertaining the paths of transferability of occupations in accordance with the need for the minimization of retraining. It is by this criterion then, if we are willing to accept it, that the competing methods must be judged.

The above statement implies a mathematical function which may be minimized (or maximized, according to its set-up), a matter which is at once both the delight and the despair of the statistician. Suffice it to say, at this stage, that the best elements of all the methods likely will yield a better result than blind adherence to any one alone. In addition it must be pointed out that the statistically most elaborate system, the WMC system, was worked out primarily for the list of occupations job-analyzed in the course of the field observations which were paralleled by the development of the Occupational Dictionary. Only a very few professions were so analyzed. The implications of this will need to be thought through in arriving at a conclusion as to the potential values of this interesting new "measurement" device. Only a complete analysis of the entire problem, including the method of "bridging the gap between man and job"¹⁰ (the coding, Hollerith sorting, and placement philosophy and practice) will reveal what elements, if any, are worthy of adoption. This statement would emphasize highly¹¹ also capacities and training (or education) to the point of specially mentioning them as well as skills.

It has been suggested that the study of occupations is a new social science.¹² Statisticians have long known that occupation, correctly ascertained—which the United States Census does not even pretend to do—is probably the most basic classification of mankind; of greater human concern, generally speaking, than race, color, sex, marital condition, education, age or place

¹⁰ One simple solution to this problem is to say, "This job requires a man to grind *valves*; this man can grind *valves*."

¹¹ Shartle, C. L. *et al* "Establishing Families of Occupations." *Vocational Guidance Magazine*, Vol. XXI (1944), 405-414.

¹² Kitson, H. D. "Occupationology—A New Science" *Occupations*, XXII (1944), 447-448.

of residence. These categories have the force of tradition behind them; but they are not as basic, psychologically, sociologically or economically as is vocation. Economically all other human values directly or indirectly stem out of or are obligated to vocation. In correctly classifying occupations, then, we are making possible not only accurate studies of this all-important institution but also are laying the foundation for gaining a control of such matters as guidance, promotion, transfer, and the conservation of human talent generally. A statement about one aspect of the matter will make clear the importance of the matter. If a worker before changing his occupation knew with a high degree of certainty whether a proposed change likely (i.e., on the average) would result in greater opportunity for him, or the converse, much human loss and unhappiness could be prevented. Again, assuming a comparable development in human traits, if one knew what was the actual distribution of wages in two occupations for people of identical trait profiles, the present unjustified discrepancies would tend, without artificial restraints, shortly to disappear. And if, finally, a similar distribution of happiness-scores (on a test of satisfactions, often called morale in the Army and industry) could also be attached to the individual profiles in the above table any individual knowing his own profile could choose his occupation so as to meet his need, whether for a greater income, greater happiness or satisfaction, or both.

The lower the prestige of one's occupation, in a scale of prestige values, the more a man expects, normally, to get his satisfactions outside his work. In the professions one's work frequently is both one's vocation and one's avocation and recreation. To maximize the individual's salary and job-originated satisfactions is one of the crucial steps in the maximization of manpower.

REFERENCES

1. Davis, Edwin W. "A Functional Pattern Technique for Classification of Jobs." *Contributions to Education* No. 844. New York: Bureau of Publications, Teachers College, 1942, p. 128.

2. Carter, Harold D. "Vocational Interests and Job Orientation," *Applied Psychology Monographs*, Vol. II (1944), 85.
3. Dvorak, Beatrice Jeanne. "Differential Occupational Ability Patterns." *University of Minnesota Employment Stabilization Research Institute*, Vol. III (1935), 46.
4. Paterson, Donald G, Geiken, C. d'A, and Hahn, M. E. *The Minnesota Occupational Rating Scales and Counselling Profile*. Chicago: Science Research Associates, 1941, p 133.
5. Trabue, M. R. "Occupational Ability Patterns." *Personnel Journal*, Vol. XI (1933), 544-551.

TESTING FOR ADMINISTRATIVE AND SUPERVISORY POSITIONS

MILTON MANDELL

United States Civil Service Commission, Washington, D. C.

THE field of testing for administrative and supervisory positions is one which is approached by many psychometricians with a feeling of defeatism. A number of studies have been made to determine which testing methods will improve the selection of administrators and supervisors but, in general, psychometricians have tended to stay away from this phase of testing. The reasons for this situation are well summarized in Dr. L. L. Thurstone's statement:

The intellectual and temperamental qualities that insure success in administrative work are probably more complex than almost any other group of abilities that can be thought of. Psychologists who investigate fundamental human traits would undoubtedly seek to investigate first those traits which can be assumed to be less complex.¹

Although the complexity of the task of experimenting in this field is recognized, there still exists in industry, in the Army and Navy, and in government, the problem of selecting persons for supervisory and administrative positions. With a labor force, both civilian and military, of about 60,000,000 in the U. S., one can roughly estimate the existence of about 2,000,000 supervisory or administrative positions. There has been little study of the validity of present selection methods for these positions, but the great interest of high administrative officials in improved selection devices is some indication of the lack of values in techniques in use at the present time.

The following are general definitions of supervisory and administrative positions. By a supervisory position is meant

¹L. L. Thurstone *A Factorial Study of Perception* University of Chicago Press, 1944

one which involves responsibility for the working conduct of, and the quality and quantity of work produced by, one or more subordinates. By an administrative position is meant one which involves extensive responsibilities for planning, organizing, directing, staffing, budgeting, and coordinating the work of an organization or part of an organization. The administrator in some cases performs no supervisory duties since he may act only in an advisory capacity, but he usually has the functions of a supervisor in addition to his administrative duties.

Supervisory positions can generally be classified on the basis of three factors: (a) the number of employees supervised, (b) the nature of the work or the occupation supervised, and (c) the supervisor's level in the organization. It is reasonable to present the hypothesis that the skills and abilities required to supervise a few employees are different in degree, and perhaps in kind, from those required to supervise hundreds of employees. For example, the supervisor who likes to check carefully and in detail the work of a few subordinates may be successful, but this same action would probably make him a bottle-neck if he applied the same method to the supervision of a hundred employees. Observation indicates that this method is as much a trait of the individual as it is a function of the position that he occupies.

The nature of the work or occupation supervised is an important distinction between supervisory positions. The methods required to supervise a gang of ditch diggers successfully differ from the methods required to supervise a staff of skilled psychometricians. Some supervisors who have transferred from organizations where strict disciplinary methods were in use have failed miserably in attempting to apply the same methods on different occupational groups or in different work situations.

The trend in the field of administration has been to consider administration as an occupation *per se* and to emphasize the common elements among such positions. This trend has been especially strong in the field of public administration, it is noticeable in Army and Navy administration, and it is evident in the field of business administration.

Those opposing this trend claim that knowledge of the business or subject-matter of the organization that is administered is also important. Agreeing for the moment that administrative skill is more important than subject-matter knowledge, however, one can still classify administrative jobs into a number of categories on the basis of other considerations. These categories can be considered to be: (a) positions whose problems are of an international or national rather than local character; (b) positions involving the "selling" of the goals and objectives of the organization to an indifferent or even hostile clientele; (c) positions of an advisory rather than operating character; (d) positions requiring the ability to administer large and complex functions; (e) positions involving swiftly moving problems as contrasted with situations where speed is not so essential; and (f) positions involving the administration of a large organization with hundreds of thousands of employees as distinguished from a small organization with a few hundred employees.²

The above categories have been presented on the assumption that some of the skills and abilities necessary for successful performance vary according to these categories. Actually, these classifications are based on job analysis and it may be found that, in terms of testing, other categories may appear to be more significant.³

Qualifications

Before discussing actual results obtained in testing for supervisory and administrative positions, it may be worth while to present various opinions on the qualities necessary for successful performance in these positions. Dr. W. V. Bingham lists a large number of such qualities: the administrator doesn't

² These differences among supervisory and administrative positions, which are intended to be illustrative, indicate that thorough job analysis is extremely essential in testing work in this area and that the psychometrician should probably have the assistance of a person trained in administration to help him identify these differences.

³ Cf. Alexander H. Leighton *The Governing of Men* Princeton: Princeton University Press, 1945. Leighton, who is a psychiatrist, in his study of a War Relocation Authority center, divided administrators into two groups, "people-minded" versus "stereotype-minded." His basic distinction between these two groups is based on whether the administrator puts emphasis on human values or on adherence to regulations.

go off "half-cocked," makes consistent decisions, has well-thought out policies, obtains real staff participation in the formulation of policies, builds a team, has ideas, listens to new ideas, is not an egotist, gives recognition for good work, delegates responsibility, sees his staff, has good timing in making decisions, and knows outside conditions.⁴ Johnson O'Connor reports that successful executives have a large vocabulary, a wide range of aptitudes, an objective personality, an accounting aptitude, and an aptitude for their first position.⁵

Schell lists as the outstanding requirements for an executive an innate interest in and affection for people, an outstanding personality, and a scientific mind.⁶ Ordway Tead, who has contributed extensively to the theory and practice of administration, has suggested the following qualities as desirable for leaders: physical and nervous energy, a sense of purpose and direction, enthusiasm, friendliness and affection, integrity, technical mastery, decisiveness, intelligence in many directions, and teaching skill.⁷ In a pamphlet published in 1923, the American Management Association recommended the following qualifications and values in selecting supervisors: personal—30%; mental—45%; moral—15%; and physical—10%.⁸ Cleeton and Mason offer as their criterion of executive ability "above average ability in a large number of qualities which can be rated or measured."⁹

It is apparent from this listing of qualities needed by administrators and supervisors that major stress is placed on aspects of personality. In addition, however, a number of other abilities are considered essential. While few would require that the administrator or supervisor should have the highest mental ability within the group that he heads, superior mental ability

⁴ W. V. Bingham *Administrative Ability*. Washington, D. C.: Society for Personnel Administration, April, 1939.

⁵ Johnson O'Connor *Characteristics of Successful Executives*. Hoboken. Stevens Institute of Technology, 1932.

⁶ Erwin H. Schell *The Technique of Executive Control*. New York: McGraw-Hill, 1930.

⁷ Ordway Tead *The Art of Leadership*. New York: Whittlesey House, 1935.

⁸ *Selecting the Supervisory Forces*. New York: American Management Association, 1923.

⁹ Glen U. Cleeton and Charles W. Mason *Executive Ability. Its Discovery and Development*. Yellow Springs: Antioch Press, 1934, p. 12.

is still considered very important.¹⁰ Part of the differences of opinion on this subject may be accounted for by the failure to relate the mental ability required to the group being supervised. The mental ability required of a foreman of laborers is obviously lower than that required of a Construction Superintendent but in relation to the mental ability of his laborers, the foreman should rank high.¹¹

The administrator and supervisor must be able to understand organization. Some technically minded persons are constantly frustrated by administrative procedures and regulations when they are promoted to supervisory positions, while others acquire a knowledge of how to get things done which is basic to success, especially in a large organization. Without some evidence it is impossible to state what factors make for success in this particular phase of administrative work, yet one can speculate that this ability is non-intellectual but related to the interests and personality of the person.¹²

Use of Paper-and-Pencil Tests

1. *Interest Inventories.* From the published studies, there is a definite indication that the measurement of interests can contribute substantially to the selection of administrative and supervisory personnel.¹³ The major problem is in determining what norms to use and which special interests, in addition to social interest or interest in people, are desirable for the particular group of positions being studied.

In his study of Federal government administrators, Thur-

¹⁰ Cf. Forrest A. Kingsbury "Psychological Tests for Executives." *Personnel*, IX (1930), 125-126

¹¹ Administrative difficulties have occurred frequently because of changes among non-supervisory workers, which resulted in a higher level of mental ability, while the supervisors, relatively unchanged, have in many cases been unable to handle these workers of superior ability. This situation was noticeable in the early 1930's when, because of economic reasons, a large number of able persons accepted employment in low-grade positions.

¹² E. K. Strong, Jr., in an unpublished study prepared in cooperation with the Committee on Public Administration of the Social Science Research Council, found that the administrators in technical fields such as law, engineering, medicine, and accounting do not generally rate their interests in the top bracket in their own profession but have broader interests, mainly in the field of dealing with persons. A part of this study has been published: E. K. Strong, Jr. "Interests of Public Administrators." *Public Personnel Review*, VI (1945), 166-173

¹³ F. H. Achard and Florence H. Clarke "You Can Measure the Probability of Success as a Supervisor" *Personnel*, XXI (1945), 355.

stone found that the Social Scale of the *Allport-Vernon Scale of Values* differentiated among his population better than any other measuring device he used.¹⁴ The Theoretical Scale of the *Scale of Values* also differentiated positively and significantly. The Commercial Interests Scale of *Thurstone's Vocational Interest Schedule* and the Religious Scale of the *Allport-Vernon* also differentiated significantly but negatively.

Achard and Clarke, in their study of 300 supervisors in the Consolidated Edison Company of New York, prepared a special rating scale of the *Strong Vocational Interest Blank for Men (Revised)* and obtained satisfactory results with four different types of supervisors. They state that "it was the best single all-around indicator of supervisory ability."¹⁵

Strong has prepared an occupational interest scale for public administrators based on his work with the Forest Service and the Committee on Public Administration. The U. S. Civil Service Commission, in its test development work in connection with the administrative intern program, has obtained promising results with the *Kuder Preference Record*.

The early studies in this field and the recent investigations by Strong, Thurstone, and Achard and Clarke indicate that studies of interest, if designed for the particular organizational situation, can contribute significantly to administrative and supervisory selection programs. In using interest inventories, it will probably be found that the appropriate critical scores vary significantly among different types of organizations and at different levels of the organization's hierarchy.

2. *Personality Inventories.* These inventories can be divided into two groups; namely, those of the omnibus type which furnish several measurements for the individual, such as the Bernreuter, and those furnishing a single score, such as Laird's scale on extroversion and introversion. While everyone is agreed that, generally speaking, the personality of the administrator or supervisor is probably the most significant single factor in contributing to successful performance, the information now available would seem to indicate that the

¹⁴ *Op cit.*, pp 142-145

¹⁵ *Op cit.*, p 362.

present personality inventories, either for reasons intrinsic or external to them, are only slightly useful for this testing purpose and are substantially less so than the interest inventories. To what extent this result may be produced by the "fudging" of responses is not known.¹⁶ It would seem that a sophisticated group of supervisors and administrators in a competitive situation would be able to "beat" these tests.

Achard and Clarke obtained promising results with the Bernreuter, scaled in accordance with their own methods.¹⁷ Their scale differentiated the good from the poor supervisors in each of the four groups they were studying better than any other test they used except the interest inventory.

Beckman and Levine obtained favorable results with the Allport's A-S test on a group of supervisory employees of the City of Cincinnati.¹⁸ In a monograph on his work with several industrial firms, Hersey reports the use of personality inventories of the omnibus type for supervisory selection. Kingsbury comments in regard to the use of extroversion-introversion tests for this purpose that "since the concepts of introversion and extroversion are so ambiguous, and the various tests intercorrelate so poorly, further analysis and experiment are needed before tests of this type can be accepted with confidence for this purpose."¹⁹ Personality inventories have frequently been used, with some promising results, for comparing leaders with non-leaders in educational institutions. Young and Cooper found, for example, in their study of elementary school children in the 5th through the 8th grades, that the leaders could be characterized as self-sufficient extroverts while neither physical nor mental characteristics or interests differentiated the leaders from the non-leaders.²⁰

Basing a conclusion on the studies made, it is evident that

¹⁶ Jurgensen's use of paired-comparison and rank-order techniques in his experimental form of a personality inventory may contribute to a substantial reduction of this factor. See C. E. Jurgensen, "Report on the Classifications Inventory," *Journal of Applied Psychology*, XXVIII (1944), 445-460.

¹⁷ *Op. cit.*, p. 362.

¹⁸ "Selecting Executives: An Evaluation of Three Tests." *Personnel Journal*, VIII (1930), 415-420.

¹⁹ *Op. cit.*, p. 127.

²⁰ "Some Factors Associated with Popularity" *Journal of Educational Psychology*, XXXV (1944), 513-535.

the present personality inventories can contribute only slightly to the selection of administrators and supervisors and that this contribution will be further reduced by the fact that candidates may "fudge" in a competitive situation. The factors that these inventories attempt to measure are so important for successful performance that much further work in this area is definitely needed.

3. *Mental Abilities.* The contribution of this type of test to predictions of supervisory success runs the gamut from extreme importance to only slight or no importance. This result is related to the previous discussion where the opinion was expressed that one cannot treat these positions as though they were identical, and the difference in the results obtained is undoubtedly the consequence of differences in job-content plus biases among raters in the organizations being studied.

The results of the Army's officer selection program indicate a fairly high positive correlation between the successful completion of officer candidate schools and general classification test scores. Thurstone in his study of Federal administrators found that the linguistics section of the *American Council on Education Psychological Examination* made a significant and positive differentiation between the better and the poorer administrators. Shuman in his study of 99 foremen obtained a correlation of $+ .39 \pm .07$ between the *Otis Q. S. Mental Ability Beta* and performance ratings.²¹ Uhrbrock and Richardson, in their thorough study of factory supervisors, found mental ability items useful for differentiating between good and poor supervisors.²² Achard and Clarke found that the *Otis Self-Administering Test of Mental Ability, Higher Examination*, was of value for each of the four groups of supervisors studied.

The above brief summary indicates that a mental ability test should be included in testing programs for supervisory and administrative positions. The weight given to the test should differ, however, depending upon the type of positions being studied.

²¹ "The Value of Aptitude Tests for Factory Workers in the Aircraft Engine and Propeller Industries," *Journal of Applied Psychology*, XXIX (1945), 159.

²² "Item Analysis, The Basis for Constructing a Test for Forecasting Supervisory Ability" *Personnel Journal*, XII (1933), 141-154.

4. *Special Types of Paper-and-Pencil Tests.* Mechanical aptitude tests have frequently been used, and often with satisfactory results, for the selection of factory supervisors. But the reason for this successful use has not been fully explained. The explanation might lie in the area of the relationship between aptitude and interests, with the successful factory supervisor being a person whose aptitude and interests are directed towards shop and mechanical situations. Whatever the reason may be, however, these tests, especially the *Bennett Mechanical Comprehension Test*, have proved useful.²³

In preliminary studies of the U. S. Forest Service, the *Interpretation of Data Test* of the Progressive Education Association has given some promising results for the selection of administrators. In various experimental studies of this test on administrative personnel, the candidates have expressed a high opinion of its value.

Uhrbrock and Richardson, in their previously cited study, found that test items on the policies and organization of the company in which the supervisors were employed made a significant contribution to the selection of supervisors. It would be our hypothesis that this type of test has value in selecting administrative or supervisory personnel from among technicians and mechanics because it would identify those persons whose interests lie beyond the technical parts of the positions they are occupying.

A device sometimes used for the selection of supervisors is the multiple-choice or true-false form for measuring the ability of a candidate to get the right answer to questions on supervisory situations. Quentin W. File's test on *How Supervise?* is one form of this type of test. Hersey has also used a similar type of test. It would seem that this type of test measures only the logical factor, and not the emotional factor, in human behavior. In other words, a supervisor may know exactly how to act in a given situation but his emotions may lead to a different procedure. This statement is not meant to deny that this type of test has value but that its primary purpose should

²³ See the previously cited studies of Shuman and Achard and Clarke

probably be to determine the need for training. Its value for selection purposes has not as yet been determined.

Thurstone in his study of Federal administrators obtained significant differences between the good and poor groups on three special types of paper-and-pencil tests, Estimating, Classification, and Gottschaldt figures. The *Estimating Test* requires the candidate, by using logical methods rather than knowledge, to arrive at a statistic. For example, a candidate might be asked the number of telephones in use for non-commercial purposes in the United States for the year 1941. It would be highly improbable that any candidate would know the answer to this question. He would have to rely on estimating to arrive at his answer. In the *Classification Test*, the candidate is asked to sort a number of cards, each containing the name of a prominent individual, into as many groups as he wishes. Thurstone found that the better administrator used fewer categories. The *Gottschaldt Figures Test* is in multiple-choice form and requires the candidate to identify in which one of five large, complicated diagrams there exists a smaller figure. The results obtained by Thurstone justify further study of the value of these tests for selecting administrative and supervisory personnel.

Oral Interviews

At some stage in the selection of supervisory and administrative personnel, an interview is practically always used. But there is no evidence available as to its value. The Army Service Forces is now engaged in a program which indicates, in at least tentative form, that by refining the questions and rating forms used, moderate validity is obtained for this testing device.

The British Army has been using what might be called a group interview in which the candidates are observed in their discussion of a subject selected either by or for them. From observation of this type of interview it would seem that this method, when properly administered, offers valuable information for administrative and supervisory positions.

One of the most common difficulties with the administration

of an oral interview is the need for careful selection and training of raters. Quite often the raters are selected on a chance basis rather than because of their technical competency. Too frequently, also, the training of the raters before the interview has been inadequate in terms of the complexity of the task assigned to them.

Ratings in Training Courses

In those organizations which conduct training programs for administrative and supervisory personnel, objective ratings on performance in these programs should furnish valuable additional information for selection for promotion. The standard conditions under which these training programs can be given would be a definite factor in helping to make these ratings valuable for this purpose. In the promotion examination program adopted recently for shop supervisory positions in Navy field establishments, a rating in the work improvement program is given a weight in determining rank on the eligible register. This examining program is under the direction of the U. S. Civil Service Commission, and includes, in addition to the work improvement program rating, a written test on administration, an oral interview, an evaluation of experience and a performance rating.

Evaluation of Biographical Data

Public personnel agencies use extensively the evaluation of a candidate's background, in terms of experience and training, as a method for rating applicants for administrative and supervisory positions. To the best of our knowledge, there is no information available as to the validity of this test for this type of position. If this testing method is based on much more extensive information than is usually contained on an application blank, it should provide a sound basis for selection if, in addition, improvements can be devised in the method for translating this information into ratings. Intensive study of the validity of this testing method is needed.

Basic to the success of any study of the value of tests is the preparation of a valid criterion. This problem is especially

complex in evaluating administrative and supervisory positions. Practically speaking, in devising tests for one organization, an attempt at agreement within the organization should be made as to the standards to be used in rating. This recognizes that different standards in use by another organization will invalidate the results obtained in the first instance. It can be pointed out that in the most successful studies in this field as much time has been spent in obtaining good ratings as on the testing program itself.

Conclusions

The conclusions we would draw from this summary of trends in testing for administrative and supervisory positions are:

(1) The testing program should include a mental ability test and an interest inventory.

(2) Further study is needed of the value of personality inventories, tests of company policies and organization, and special types of tests such as the *Interpretation of Data* and the *Thurstone Estimating Test*.

(3) The value of ratings in training courses and on biographical data should be explored.

(4) Precise analysis of the jobs being studied is an absolute need in order to identify homogeneous sub-groups.

(5) Differences in the value of particular tests should be expected for different supervisory and administrative jobs.

(6) The evident practical urge to include oral interviews makes necessary further improvements in this testing technique.

ARMY GENERAL CLASSIFICATION TEST SCORES FOR CIVILIAN OCCUPATIONS

THOMAS W HARRELL¹

University of Illinois

and

MARGARET S HARRELL

FOR more than twenty years research workers have been referring to data on intelligence and occupation based on the *Army Alpha* scores and the civilian occupations of enlisted men in World War I.² The present study deals with similar material. It reports the *Army General Classification Test (GCT)* scores of 18,782 white enlisted men of the Army Air Forces Air Service Command³ distributed according to their previous civilian occupation. Means and medians and standard deviations are reported for the 74 occupations for which there were samples of sufficient size to be of significance.

The data, more applicable to occupations of today than are the data presented in earlier studies, supplement present knowledge concerning ability levels for occupational groups and are of value in educational and vocational counseling of civilians and especially, at this time, of discharged soldiers.

There has been much discussion as to how well the *Army Alpha* sample represented the general population. At the present time it is impossible to decide that issue with respect to this sample. It is possible that the averages among the professional occupations are too low since conceivably many

¹ On leave, Major, U. S. Army Air Corps, Director, Manning Section, Headquarters 15th Air Force.

² Yerkes, R. M. "Psychological Examining in the U. S. Army." *Memoirs of the National Academy of Sciences*, 1921.

Fryer, Douglas "Occupational-Intelligence Standards" *School and Society*, XVI (1922), 273-277

Bingham, Walter V. *Aptitudes and Aptitude Testing*. New York: Harper Bros., 1937, pp. 44-59.

³ Data were furnished by Lt. Col. R. W. Faubion and Lt. Col. J. L. Webster.

of the best men in the profession would have been officer material. It is more likely that the averages among the lowest scoring occupations are too high since for a while the Army Air Forces were receiving a smaller percentage of enlisted men with low scores on the *GCT* than were the Army Ground and the Army Service Forces.

Scores were obtained from Informational Rosters made up from Soldier's Qualification Cards and other sources. Only those cases were included where the roster clearly indicated the previous civilian occupation of the soldier by job title. The job titles in the following tables are those given and described in Army Regulation 615-26, Index and Specifications for Civilian and Military Occupational Specialists, or in the U. S. Employment Service Dictionary of Occupational Titles. In a few instances, however, a general job title has been used in this study instead of the more specific breakdowns of the job described in A.R. 615-26, i.e., Engineer includes mechanical, civil and mining engineers. These general job titles are Engineer, Laboratory Assistant, Inspector, Musician, Foreman, Electrician, Assembler and Welder, Manager, and Miscellaneous. Miscellaneous is a general title which includes managers of various business establishments, i.e., moving picture theaters, bowling alleys, etc., but excludes managers of retail stores. Manager, Retail Store includes managers of both chain and independent retail stores. Mechanic includes all kinds of mechanics except air-plane and automobile mechanics.

The desired maximum size of sample for a single occupation was 500 cases, and when that maximum was reached further tabulation of the occupation was discontinued except in a few instances where the range of scores was great. For the majority of the occupations 500 cases were not available. It was originally planned to omit all occupations with less than 100 cases; some have been included, however, as they involve several professional groups in which there is considerable interest.

In order to provide a rough check on the reliability of the size of the sample, the scores for each occupation were tabulated in two distributions, A and B, of approximately the same

TABLE 1

Mean and Median GGT Standard Scores, Standard Deviations and Range of Scores of 18,782 AAF White Enlisted Men by Civilian Occupation

Occupation	N	M	Median	Standard deviation	Range
Accountant	172	128.1	128.1	11.7	94-157
Lawyer	94	127.6	126.8	10.9	96-157
Engineer	39	126.6	125.8	11.7	100-151
Public Relations Man	42	126.0	125.5	11.4	100-149
Auditor	62	125.9	125.5	11.2	98-151
Chemist	21	124.8	124.5	13.8	102-153
Reporter	45	124.5	125.7	11.7	100-157
Chief Clerk	165	124.2	124.5	11.7	88-153
Teacher	256	122.8	123.7	12.8	76-155
Draftsman	153	122.0	121.7	12.8	74-155
Stenographer	147	121.0	121.4	12.5	66-151
Pharmacist	58	120.5	124.0	15.2	76-149
Tabulating Machine Operator	140	120.1	119.8	13.3	80-151
Bookkeeper	272	120.0	119.7	13.1	70-157
Manager, Sales	42	119.0	120.7	11.5	90-137
Purchasing Agent	98	118.7	119.2	12.9	82-153
Manager, Production	34	118.1	117.0	16.0	82-153
Photographer	95	117.6	119.8	13.9	66-147
Clerk, General	496	117.5	117.9	13.0	68-155
Clerk-Typist	468	116.8	117.3	12.0	80-147
Manager, Miscellaneous	235	116.0	117.5	14.8	60-151
Installer-Repairman, Tel & Tel	96	115.8	116.8	13.1	76-149
Cashier	111	115.8	116.8	11.9	80-145
Instrument Repairman	47	115.5	115.8	11.9	82-141
Radio Repairman	267	115.3	116.5	14.5	56-151
Printer, Job Pressman, Lithographic Pressman	132	115.1	116.7	14.3	60-149
Salesman	494	115.1	116.2	15.7	60-153
Artist	48	114.9	115.4	11.2	82-139
Manager, Retail Store	420	114.0	116.2	15.7	52-151
Laboratory Assistant	128	113.4	114.0	14.6	76-147
Tool Maker	60	112.5	111.6	12.5	76-143
Inspector	358	112.3	113.1	15.7	54-147
Stock Clerk	490	111.8	113.0	16.3	54-151
Receiving and Shipping Clerk	486	111.3	113.4	16.4	58-155
Musician	157	110.9	112.8	15.9	56-147
Machinist	456	110.1	110.8	16.1	38-153
Foreman	298	109.8	111.4	16.7	60-151
Watchmaker	56	109.8	113.0	14.7	68-147
Airplane Mechanic	235	109.3	110.5	14.9	66-147
Sales Clerk	492	109.2	110.4	16.3	42-149
Electrician	289	109.0	110.6	15.2	64-149
Lathe Operator	172	108.5	109.4	15.5	64-147
Receiving & Shipping Checker	281	107.6	108.9	15.8	52-151
Sheet Metal Worker	498	107.5	108.1	15.3	62-153
Lineman, Power and Tel & Tel	77	107.1	108.8	15.5	70-133

TABLE 1 (Continued)

Occupation	N	M	Median	Standard deviation	Range
Assembler	498	106.3	106.6	14.6	48-145
Mechanic	421	106.3	108.3	16.0	60-155
Machine Operator	486	104.8	105.7	17.1	42-151
Auto Serviceman	539	104.2	105.9	16.7	30-141
Riveter	239	104.1	105.3	15.1	50-141
Cabinetmaker	48	103.5	104.7	15.9	66-127
Upholsterer	59	103.3	105.8	14.5	68-131
Butcher	259	102.9	104.8	17.1	42-147
Plumber	128	102.7	104.8	16.0	56-139
Bartender	98	102.2	105.0	16.6	56-137
Carpenter, Construction	451	102.1	104.1	19.5	42-147
Pipe Fitter	72	101.9	105.2	18.0	56-139
Welder	493	101.8	103.6	16.1	48-147
Auto Mechanic	466	101.3	101.8	17.0	48-151
Molder	79	101.1	105.5	20.2	48-137
Chauffeur	194	100.8	103.0	18.4	46-143
Tractor Driver	354	99.5	101.6	19.1	42-147
Painter, General	440	98.3	100.1	18.7	38-147
Crane Hoist Operator	99	97.9	99.1	16.6	58-147
Cook and Baker	436	97.2	99.5	20.8	20-147
Weaver	56	97.0	97.3	17.7	50-135
Truck Driver	817	96.2	97.8	19.7	16-149
Laborer	856	95.8	97.7	20.1	26-145
Barber	103	95.3	98.1	20.5	42-141
Lumberjack	59	94.7	96.5	19.8	46-137
Farmer	700	92.7	93.4	21.8	24-147
Farmhand	817	91.4	94.0	20.7	24-141
Miner	156	90.6	92.0	20.1	42-139
Teamster	77	87.7	89.0	19.6	46-145

size. The difference between the means of A and B for each of the 48 occupations with 100 or more cases ranged from 0.1 to 5.1 and averaged 1.7. (The median difference was 1.3.) For the 26 occupations with less than 100 cases, the difference between the means of groups A and B ranged from 0.4 to 11.7 and averaged 3.7. (The median difference was 3.0.)

Table 1 gives the mean, median, standard deviation and range for 74 occupations. The means and medians were calculated from distributions grouping scores by intervals of two. The standard deviations were calculated from distributions grouped to provide from 12 to 18 class intervals.

Table 2 gives the percentage distribution of each occupation. It shows, as have other similar studies, the great overlapping of scores between occupations. Even Teamster, the

TABLE 2
Percentage Distribution of GCT Standard Scores by Civilian Occupation of 18,782 AAF White Enlisted Men

Accountant	Lawyer	Engineer	Pub. Rel Man	Auditor	Chemist	Reporter	Chief Clerk	Teacher	Draftsman	Stenographer	Pharmacist	Tab. Mach. Oper.	Bookkeeper	Sales Mgr.	Purch Agent	Prod. Mgr	Photographer	Clerk, Genl	Clerk-Typist	Misc	Inst.-Repair, Tel & Tel	Cashier	Instrument Repairman	Radio Repair	Printer	Salesman	Artist	Mgr. Retail Store	Lab Asst.	Tool Maker	Inspector	Stock Clerk	R/S Clerk	Musician	Machinist	Foreman		
150-159	3	2	8	2	9	4	1	1	2	1	4	1	2	3	1	3	3	1	3	4	*	3	1	2	2	4	2	*	4	3	2	1	1	1	2	2	150-159	
140-149	13	14	5	12	10	2	8	15	8	17	5	7	14	17	16	20	10	12	9	11	7	3	2	2	8	13	2	2	4	8	2	8	9	6	5	2	140-149	
130-139	25	17	15	17	17	5	18	16	15	37	17	13	14	38	28	18	37	28	30	29	33	32	21	31	27	24	2	2	8	2	12	8	9	6	5	130-139		
120-129	31	47	46	36	38	29	35	37	31	34	41	29	29	38	28	18	37	28	30	29	33	32	21	31	27	24	2	13	20	23	17	20	18	25	21	22	120-129	
110-119	22	15	18	26	24	29	30	26	24	28	41	26	32	19	30	27	24	31	31	24	26	30	32	29	27	28	50	30	29	26	28	31	27	24	25	110-119		
100-109	4	4	8	9	5	14	9	6	8	18	11	5	18	13	19	14	20	18	16	19	20	18	20	23	18	19	9	17	20	30	23	19	18	25	23	100-109		
90-99	2	1	1	1	1	4	1	3	1	4	2	4	6	7	5	3	4	7	6	7	9	6	9	7	10	7	4	8	13	8	11	14	10	13	12	10	90-99	
80-89	2	1	1	1	1	1	1	2	0	1	2	2	1	3	3	6	2	2	2	4	0	1	2	3	4	3	3	4	4	3	5	7	6	6	8	80-89		
70-79	2	1	1	1	1	1	1	1	1	1	2	2	1	3	3	9	1	1	1	0	1	1	2	2	1	0	3	6	2	2	2	3	2	4	3	70-79		
60-69	2	1	1	1	1	1	1	1	1	1	2	2	1	3	3	6	1	1	1	0	1	1	1	1	2	3	3	6	2	2	5	5	7	2	1	2	60-69	
50-59	2	1	1	1	1	1	1	1	1	1	2	2	1	3	3	9	1	1	1	1	1	1	1	1	2	3	3	6	2	2	5	5	7	2	1	1	50-59	
40-49	2	1	1	1	1	1	1	1	1	1	2	2	1	3	3	9	1	1	1	1	1	1	1	1	2	3	3	6	2	2	5	5	7	2	1	1	0	40-49
30-39	2	1	1	1	1	1	1	1	1	1	2	2	1	3	3	9	1	1	1	1	1	1	1	1	2	3	3	6	2	2	5	5	7	2	1	1	0	30-39
20-29	2	1	1	1	1	1	1	1	1	1	2	2	1	3	3	9	1	1	1	1	1	1	1	1	2	3	3	6	2	2	5	5	7	2	1	1	0	20-29
10-19	2	1	1	1	1	1	1	1	1	1	2	2	1	3	3	9	1	1	1	1	1	1	1	1	2	3	3	6	2	2	5	5	7	2	1	1	0	10-19

* Fewer than 0.5% cases

TABLE 2 (Continued)

150-159	140-149	130-139	120-129	110-119	100-109	90-99	80-89	70-79	60-69	50-59	40-49	30-39	20-29	10-19
Watchmaker	2	1	2	5	6	5	4	4	4	1	0	*		
Arpl Mech	2	1	2	5	6	5	4	4	4	1	0	*		
Sales Clerk	2	1	2	5	6	5	4	4	4	1	0	*		
Electrician	2	2	2	5	6	5	4	4	4	1	0	*		
Lathe Operator	2	2	2	5	6	5	4	4	4	1	0	*		
R/S Checker	2	2	2	5	6	5	4	4	4	1	0	*		
Sheet Metal Worker	2	2	2	5	6	5	4	4	4	1	0	*		
Lineman	2	2	2	5	6	5	4	4	4	1	0	*		
Assembler	2	2	2	5	6	5	4	4	4	1	0	*		
Mechanic	2	2	2	5	6	5	4	4	4	1	0	*		
Machine Oper	2	2	2	5	6	5	4	4	4	1	0	*		
Auto Service	2	2	2	5	6	5	4	4	4	1	0	*		
Riveter	2	2	2	5	6	5	4	4	4	1	0	*		
Cabinetmaker	2	2	2	5	6	5	4	4	4	1	0	*		
Upholsterer	2	2	2	5	6	5	4	4	4	1	0	*		
Butcher	2	2	2	5	6	5	4	4	4	1	0	*		
Plumber	2	2	2	5	6	5	4	4	4	1	0	*		
Bar tender	2	2	2	5	6	5	4	4	4	1	0	*		
Carpenter	2	2	2	5	6	5	4	4	4	1	0	*		
Pipe Fitter	2	2	2	5	6	5	4	4	4	1	0	*		
Welder	2	2	2	5	6	5	4	4	4	1	0	*		
Auto Mechanic	2	2	2	5	6	5	4	4	4	1	0	*		
Molder	2	2	2	5	6	5	4	4	4	1	0	*		
Chauffeur	2	2	2	5	6	5	4	4	4	1	0	*		
Tractor Driver	2	2	2	5	6	5	4	4	4	1	0	*		
Painter	2	2	2	5	6	5	4	4	4	1	0	*		
Crane Hoist Operator	2	2	2	5	6	5	4	4	4	1	0	*		
Cook & Baker	2	2	2	5	6	5	4	4	4	1	0	*		
Weaver	2	2	2	5	6	5	4	4	4	1	0	*		
Truck Driver	2	2	2	5	6	5	4	4	4	1	0	*		
Laborer	2	2	2	5	6	5	4	4	4	1	0	*		
Barber	2	2	2	5	6	5	4	4	4	1	0	*		
Lumberjack	2	2	2	5	6	5	4	4	4	1	0	*		
Farmer	2	2	2	5	6	5	4	4	4	1	0	*		
Farmland	2	2	2	5	6	5	4	4	4	1	0	*		
Miner	2	2	2	5	6	5	4	4	4	1	0	*		
Teamster	2	2	2	5	6	5	4	4	4	1	0	*		

* Fewer than 0.5% cases

TABLE 3

Critical Ratios between the Means of Pairs of 74 Civilian Occupations (N = 13,782 White Enlisted Men, AAF)

	Accountant	Lawyer	Engineer	Pub Rel Man	Auditor	Chemist	Reporter	Chief Clerk	Teacher	Draftsman	Stenographer	Pharmacist	Tab Mach	Oper.	Bookkeeper	Sales Mgr.	Purch Agent	Prod. Mgr.	Photographer	Clerk, Gen'l
Accountant		3	7	1	1.3	1	1	18	*											
Lawyer	3		.5	8	.9	9	1.5	23												
Engineer	7	5		2	3	5	8	12	19	2.1	26	22				29		26		
Pub Rel Man	11	8	.2		0	.3	6	.9	17	20	25	21	28		2.8		2.4			
Auditor	13	.9	.3	0		3	6	10	19	22	28	22					2.5			
Chemist	1.1	.9	5	3	3		1	.2	.6	9	12	12	1.5	15	17	19	16	2.2	24	
Reporter	18	15	8	6	.6	1	.2	.9	12	17	15	21	23	22	27	20				
Chief Clk.	*	23	12	.9	10	2	2		12	16	23	17	2.8	26		21				
Teacher			1.9	17	1.9	6	9	1.2	.6	14	11	2.0	2.5	20	27	16				
Draftsman			2.1	2.0	2.2	.9	1.2	1.6	.6		7	7	1.2	1.5	15	20	1.3	2.5		
Steno.			2.6	2.5	2.8	1.2	1.7	2.3	1.4	7		2	.6	8	1.0	14	10	19		
Pharmacist			2.2	2.1	2.2	1.2	1.5	1.7	1.1	7	2		2	.2	6	8	7	1.2	14	
Tab. Mach Op				2.8		1.5	2.1	2.8	2.0	1.2	6	.2		1	5	8	7	14	20	
Bookkeeper						1.5	2.3	2.5	1.5	.8	2	1		5	.8		3	1.5	2.5	
Sales Mgr.			2.9	2.8		1.7	2.2	2.6	2.0	1.5	1.0	.6	.5	5		.1	3	.6	8	
Purch. Agnt						1.9	2.7		27	20	14	8	8	8	1		2	.6	8	
Prod Mgr			2.6	24	25	1.6	2.0	2.1	16	13	10	.7	.7	7	3	.2		.2	2	
Photogrphr.						2.2			2.5	1.9	1.2	14	1.5	.6	6		2		.1	
Clk., Gen'l						24					14	20	2.5	.8	8	2	1			
Clk.-Typist						26					18	26		12	13	5	5	9		
Mgr Misc						28					20	28		1.5	17	7	9	13		
In.-Rep., T & T						27					20	25	27	14	16	8	9	12		
Cashier						28					21	27		1.5	17	.8	10	13		
Inst. Rep						27					27	19	22	24	14	1.5	8	.9	1.1	
Radio Rep.											24			19	22	10	14	21		
Printer											23			1.8	20	10	13	17		
Saleman											25			2.0	24	11	16	26		
Artist						29					22	26	28	17	18	10	13	1.5		
Mgr Ret. Str.														26		14	22			
Lab Asst.														26	29	16	22	29		
Tool Maker														27		18	24	29		
Inspector																20				
Stock Clerk																22				
R/S Clerk																24				
Musician																24				
Machinist																28				
Foreman																29				
Watchmaker																25				

* Where the critical ratio is 3 or more, no entry has been made in the table

TABLE 3 (Continued)

Critical Ratios between the Means of 74 Civilian Occupations (N=18,782 White Enlisted Men, AAF)

	Clerk-Typist	Mgr, Misc	Inst-Rep, Tel & Tel	Cashier	Inst Repairman	Radio Repairmen	Printer	Salesman	Artist	Mgr, Ret Store	Lab Asst	Tool Maker	Inspector	Stock Clerk	R/S Clerk	Musician	Machinist	Foreman
Chemist	2.6	2.8	2.7	2.8	2.7	*			2.9									
Reporter	*																	
Chief Clk																		
Teacher																		
Draftsman																		
Steno.						2.7												
Pharmacist	1.8	2.0	2.0	2.1	1.9	2.4	2.3	2.5	2.2									
Tab Mach Op	2.6	2.8	2.5	2.7	2.2				2.6									
Bookkeeper			2.7		2.4				2.8									
Sales Mgr.	1.2	1.5	1.4	1.5	1.4	1.9	1.8	2.0	1.7	2.6	2.6	2.7						
Purch Agnt.	1.3	1.7	1.6	1.7	1.5	2.2	2.0	2.4	1.8		2.9							
Prod, Mgr	.5	.7	.8	.8	.8	1.0	1.0	1.1	1.0	1.4	1.6	1.8	2.0	2.2	2.4	2.4	2.8	2.9
Photogrphr.	.5	.9	.9	1.0	.9	1.4	1.3	1.6	1.3	2.2	2.2	2.4						
Clk, Gen'l	9	13	1.2	1.3	1.1	2.1	1.7	2.6	1.5		2.9	2.9						
Clk-Typist		.7	.7	.8	.7	1.4	1.3	1.9	1.1	2.9	2.4	2.5						
Mgr, Misc.	.7		1	.1	.3	.5	.6	.8	6	1.6	1.6	1.9	2.9					
In-Rep., T & T	7	.1		0	1	3	4	.5	4	1.2	1.3	1.6	2.2	2.6	2.9	2.7		
Cashier	8	.1	0		1	3	4	.5	5	1.3	1.4	1.7	2.5			2.9		
Inst Rep.	.7	3	.1	.1		.1	.2	.2	.3	.8	1.0	1.3	1.7	2.0	2.2	2.1	2.9	2.9
Radio Rep	1.4	5	.3	.3	.1		.1	.2	2	1.1	1.2	1.5	2.5			2.8		
Printer	1.3	.6	.4	.4	.2	1		0	1	.8	.9	1.3	1.9	2.3	2.6	2.4		
Salesman	1.9	.8	.5	.5	.2	2	0		.1	1.1	1.2	1.5	2.6			2.9		
Artist	1.1	.6	.4	.5	.3	.2	1	1		.5	7	1.1	1.4	1.7	2.0	1.9	2.7	2.7
Mgr Ret Store	2.9	1.6	1.2	1.3	.8	1.1	.8	1.1	.5		.4	.8	1.5	2.1	2.5	2.1		
Lab Asst.	2.4	1.6	1.3	1.4	1.0	1.2	.9	1.2	7	.4		4	7	1.1	1.4	1.4	2.2	2.2
Tool Maker	2.5	1.9	1.6	1.7	1.3	1.5	1.3	1.5	1.1	.8	.4		1	.4	.7	.8	1.3	1.4
Inspector		2.9	2.2	2.5	1.7	2.5	1.9	2.6	1.4	1.5	.7	1		.5	.9	.9	2.0	2.0
Stock Clk			2.6		2.0		2.3		1.7	2.1	1.1	.4		.5		.6	1.6	1.6
R/S Clerk			2.9		2.2		2.6		2.0	2.5	1.4	.7	.9	.5		3	1.1	1.2
Musician			2.7	2.9	2.1	2.8	2.4	2.9	1.9	2.1	1.4	.8	.9	.6	.3		.5	.7
Machinist					2.9				2.7		2.2	1.3	2.0	1.6	1.1	.5		2
Foreman					2.9				2.7		2.2	1.4	2.0	1.6	1.2	.7	.2	
Watchmaker			2.8	2.5	2.6	2.2	2.6	2.3	2.5	2.0	2.0	1.5	1.1	1.2	1.0	.7	.5	1.0
Airpl Mech												2.5	1.7	2.3	2.0	1.6	1.0	.7
Sales Clk.												2.8	1.9	2.8	2.5	2.0	1.2	.9
Electric'n												2.8	1.9	2.7	2.4	2.0	1.2	1.0
Lathe Oper.												2.8	2.0	2.6	2.4	2.0	1.4	1.1
R/S Checker													2.6			2.1	2.1	1.6
Sheet Mtl Wkr.													2.9			2.4	2.5	1.9
Lineman												2.9	2.3	2.7	2.5	2.2	1.8	1.6
Assembler																		
Mechanic																		2.8
Mach. Oper																		
Auto Serv																		
Riveter																		
Cabinetmkr.																		2.8

* Where the critical ratio is 3 or more, no entry has been made in the table

TABLE 3 (Continued)

Critical Ratios between the Means of 74 Civilian Occupations (N=18,782 White Enlisted Men, AAF)

	Watchmaker	Appl. Mech	Sales Clerk	Electrician	Lathe Operator	R/S Checker	Sht. Met Worker	Lineman	Assembler	Mechanic	Machine Oper.	Auto Service	Riveter	Cabinetmaker	Upholsterer	Butcher	Plumber	Bartender
Lab. Asst	**	1.5	2.5	2.8	2.8	2.8	*	2.9										
Tool Maker	1.1	1.7	1.9	1.9	2.0	2.6	2.9	2.3										
Inspector	1.2	2.3	2.8	2.7	2.6			2.7										
Stock Clk	1.0	2.0	2.5	2.4	2.4			2.5										
R/S Clerk	.7	1.6	2.0	2.0	2.0			2.2										
Musician	.5	1.0	1.2	1.2	1.4	2.1	2.4	1.8						2.8				
Machinist	.1	.7	.9	1.0	1.1	2.1	2.5	1.6						2.7				
Foreman	0	.4	.5	.6	.8	1.6	1.9	1.3		2.8				2.5				
Watchmaker		.2	.3	.4	.6	1.0	1.1	1.0	1.7	1.7	2.4	2.7	2.6	2.1	2.4	2.9	2.9	
Appl. Mech	.2		1	.2	.5	1.3	1.5	1.1	2.6	2.4				2.3	2.8			
Sales Clk	3	1		2	.5	1.3	1.7	1.1		2.7				2.4	2.9			
Electric'n	.4	2	2		.3	1.1	1.3	1.0	2.4	2.3				2.2	2.7			
Lathe Oper	6	5	5	3		6	7	7	1.6	1.5	2.6		2.9	1.9	2.3			
R/S Checker	1.0	1.3	1.3	1.1	.6		.1	.3	1.1	1.1	2.3	2.9	2.6	1.7	2.0	2.9	2.8	
Sheet Mtl Wkr	1.1	1.5	1.7	1.3	.7	1		2	1.3	1.2	2.6		2.9	1.7	2.1		2.9	
Lineman	1.0	1.1	1.1	1.0	.7	.3	.2		4	.4	1.2	1.5	1.5	1.2	1.5	2.0	1.9	2.0
Assembler	1.7	2.6		2.4	1.6	1.1	1.3	4	0	1.5	2.2	1.9	1.2	1.5	2.7	2.3	2.3	
Mechanic	1.7	2.4	2.7	2.3	1.5	1.1	1.2	4	0	1.4	2.0	1.8	1.2	1.5	2.6	2.2	2.2	
Mach. Oper	2.4				2.6	2.3	2.6	1.2	1.5	1.4		6	.6	5	7	1.4	1.3	1.4
Auto Serv.	2.7				2.9			1.5	2.2	2.0	.6		1	3	4	1.0	1.0	1.1
Riveter	2.6				2.9	2.6	2.9	1.5	1.9	1.8	.6	1		2	.4	8	8	1.0
Cabinetmkr	2.1	2.3	2.4	2.2	1.9	1.7	1.7	1.2	1.2	1.2	.5	.3	.2		.1	2	3	.5
Upholsterer	2.4	2.8	2.9	2.7	2.3	2.0	2.1	1.5	1.5	1.5	.7	.4	.4	1		.2	3	4
Butcher	*							2.0	2.7	2.6	1.4	1.0	8	.2	.2		.1	4
Plumber	2.9				2.9			1.9	2.3	2.2	1.3	1.0	8	3	3	.1		2
Bartender	2.9				2.8	2.9		2.0	2.3	2.2	1.4	1.1	1.0	.5	4	4	2	
Carpenter								2.5			2.3	1.8	1.5	.6	.6	.6	4	1
Pipe Fitter	2.7				2.7	2.5	2.5	1.9	2.0	1.9	1.3	1.0	9	5	5	.4	3	1
Welder								2.8			2.8	2.4	1.9	7	7	9	.6	2
Auto Mech											2.7	2.2		9	1.0	1.2	.9	5
Molder					2.9	2.6	2.7	2.1	2.2	2.2	1.5	1.3	1.2	7	.7	.7	.6	.4
Chauffeur								2.9			2.6	2.3	2.0	1.0	1.1	1.2	1.0	7
Tractor Drvr.														1.6	1.8	2.3	1.8	1.4
Painter														2.1	2.4		2.6	2.1
Crane Hoist Op.														2.0	2.1	2.5	2.2	1.8
Cook and Baker														2.5	2.9			2.6
Weaver														2.9	2.8	2.0	2.1	1.8
Truck Drvr.																		
Laborer																		
Barber														2.7	2.9			2.6
Lumberjack														2.6	2.7	2.9	2.7	2.4

* Where the critical ratio is 3 or more, no entry has been made in the table

** The following are additional critical ratios for Watchmaker: Watchmaker and Prod Mgr, 2.5, and Mgr, Misc, 2.8, Tel & Tel Installer-Repairman, 2.5, Cashier, 2.6, Instrument Repair, 2.2, Radio Repair, 2.6; Printer, 2.3; Salesman, 2.5, Artist, 2.0, Mgr Retail Store, 2.0.

TABLE 3 (Continued)

Critical Ratios between the Means of 74 Civilian Occupations ($N=18,782$ White Enlisted Men, AAF)

	Carpenter	Pipe Fitter	Welder	Auto Mechanic	Molder	Chauffeur	Tractor Driver	Painter	Crane Hoist Operator	Cook & Baker	Weaver	Truck Driver	Laborer	Barber	Lumberjack	Farmer	Farmhand	Miner	Teamster
Watchmkr	* 27																		
Airpl. Mech.																			
Sales Clk																			
Electrtn																			
Lathe Op	27			29															
R/S Checker	25			26															
S. Mtl. Wkr.	25			27															
Lineman	25	19	28	21	29														
Assmblr	20			22															
Mechanic	19			22															
Mach Oper.	23	13	28	15	26														
Auto Serv	18	10	24	27	13	23				29									
Riveter	15	9	19	22	12	20				28									
Cabinetmkr.	6	5	7	9	7	10	16	21	20	25	20				27	26			
Upholsterer	6	5	7	10	7	11	18	24	21	29	21				29	27			
Butcher	6	4	9	12	7	12	23	25	23						29				
Plumber	4	3	6	9	6	10	18	26	22	21					27				
Bartender	1	1	2	5	4	7	14	21	18	26	18				26	24			
Carpenter	1	3	7	4	8	19		22		20					27				
Pipefitter	1	0	3	3	4	10	16	15	20	15	26	27	23	22					
Welder	3	0		5	3	7	18	21	19						26				
Auto Mech.	7	3	5	1	3	14	25	18	17						28	24			
Molder	4	3	3	1	1	6	11	11	16	13	21	22	19	19					
Chauffeur	8	4	7	3	1	8	16	14	22	14					23	21			
Tract Drvr.	19	10	18	14	6	8	9	8	16	10	27				19	17			
Painter	16	25	11	16	9		2	8	5	19	22	14	13						
Crane Hst Op	22	15	21	18	11	14	8	2	4	3	9	12	10	10	28				
Cook-Baker	20			16	22	16	8	4	1	8	12	8	9						
Weaver	20	15	19	17	13	14	10	5	3	1	3	5	5	7	17	22	22	29	
Truck Drvr	26			21	27	19	9	8	3		4	4	6						
Laborer	27			22			22	12	12	5	4		2	4	29				
Barber	2			28	19	23	19	14	10	8	5	4	2		2	12	18	18	25
Lumberjack	27	2	26	24	19	21	17	13	10	9	7	6	4	2		7	12	13	21
Farmer									28		17	29	12	7		12	12	21	
Farmhand											22		18	12	12			5	15
Miner										22				18	13	12	5		11
Teamster										29				25	21	21	16	11	

* Where the critical ratio is 3 or more, no entry has been made in the table.

Occupation with the lowest mean, has 2% of its cases scoring as high as the mean of Accountant, the occupation with the highest mean. On the other hand the top seven occupations have no cases as low as the mean score for Teamster. Evidently a certain minimum of intelligence is required for any one of many occupations and a man must have that much intelligence in order to function in that occupation, but a man may have high intelligence and be found in a lowly occupation because he lacks other qualifications than intelligence.

Table 3 gives the critical ratios between the means of each occupation compared with every other occupation. Where the critical ratio was 3 or more no entry has been made in the table. The table shows, as does Table 2, both the overlapping and differentiation between occupations. In general, the mean of an occupation is not significant within a range of approximately ± 10 occupations where the occupations have been arranged in descending order of means. This range would have been less if all of the occupations had had at least 100 cases. It is all the more interesting, consequently, that Chemist with 21 cases, Engineer with 39, Public Relations Man with 42 and Reporter with 45 cases have means significantly higher than those of at least 49 other occupations.

As is always found in studies of this kind, the professional and clerical administrative groups of occupations show the highest average scores. Those occupations with average scores of 120 or more are, in the order of their superiority: Accountant, Lawyer, Engineer, Public Relations Man, Auditor, Chemist, Reporter, Chief Clerk, Teacher, Draftsman, Stenographer, Pharmacist, Tabulating Machine Operator and Bookkeeper.

Since the *GCT* is a measure of ability to manipulate words, numbers and space relations, it is to be expected that those occupations with the lowest averages on the test are likewise the occupations least concerned with words, numbers or space relations. Lowest scoring occupations from low to high are Teamster, Miner, Farmhand, Farmer, Lumberjack, Barber, Laborer and Truck Driver.

MECHANICAL ABILITY, ITS NATURE AND MEASUREMENT. I. AN ANALYSIS OF THE VARIABLES EMPLOYED IN THE PRELIMINARY MINNESOTA EXPERIMENT

J. R. WITTENBORN

Yale University

A Review of the Minnesota Program

THE present study comprises a brief review of the preliminary Minnesota experiment and a factorial analysis of the variables. The purpose of this paper is to apply modern analytical techniques to the interesting problem of the nature of mechanical ability which was begun some two decades ago by the authors of the Minnesota experiment (5)

Although the Minnesota research had as its primary aim the discovery and measurement of mechanical ability, the term "mechanical ability" has never been rigidly and unambiguously defined. Since what is meant by mechanical ability has never been perfectly general, procedures for assaying the degree of mechanical ability are held to be appropriate by various investigators to varying degrees. In general, however, mechanical ability has been an expression for whatever ability or abilities are required for creditable work with tools and machinery, i.e., mechanical work. The exact nature of mechanical ability was not specified by the authors of the Minnesota study; it was instead likened by them to intelligence. It was considered to be a general function, probably comprising constituent parts, the precise identity and importance of each varying from authority to authority.

The question of the organization and composition of mechanical ability is, of course, critical for economical testing procedures. Since the publication of the Minnesota studies, new analytical techniques have been devised and employed.

It is possible that by combining the contributions of various investigators a tentative answer to the question of the organization of mechanical ability may be offered. The answer must be regarded as tentative not only because our researches to date are by no means exhaustive but also for a more positive reason. The performances which investigators have found desirable to scrutinize and to measure are determined by the requirements and the characteristics of our culture. As mechanical and other practical operations change perhaps old requirements become less important and new ones rise to importance. Different technical devices will be developed and the precise composition of that loosely defined group of abilities referred to as "mechanical" must be expected to change.

Of the several patterns of speculation concerning the composition of abilities, the theory of unique traits has been most provocative. This theory has been linked with such names as Hull, Kelley, Thurstone, Thorndike, and Woodrow. It, in general, provides two assumptions: (1) that all variability in human behavior may be expressed as a function of a limited number of independent, elemental abilities and (2) that these abilities may be discovered and suitable tests designed. This theory had been influencing psychometric research some time prior to the publication of the *Minnesota Mechanical Ability Tests*. Its full effect was not felt, however, until the invention of a practicable factor analysis by Thurstone. Before the development of effective factorial methods, a variety of expedients were employed by those who desired to make explorations in the area of human ability.

In the preliminary experiment, the literature was carefully surveyed for the purpose of selecting tests which were known to possess encouraging validity and reliability. The practical problems involved in their administration were considered. Particular effort was made in the selection to avoid including more than one test which appeared to measure the same aspect of mechanical ability. Twenty-six tests were selected, reorganized, and in some cases greatly changed with respect to instructions, conditions of administration, or length. Although the authors claimed that test individuality was emphasized in

the selection, it was possible for them to make a provisional classification of the tests under seven general headings. The headings indicated the nature of the operation which the tests appear to be measuring (5, 43-44).

I. Standard group intelligence tests

1. Army Alpha, Form 6 (group paper)
2. Otis Self-Administering Tests of Mental Ability, Higher Examination: Form A (group paper)

II. Simple motor tests

1. Tapping Test A (group paper)
2. Tapping Test B (group paper)
3. Tapping Test C (individual apparatus)
4. Steadiness of Motor Control (individual apparatus)
5. Accuracy of Movement or Tracing Paper (group paper)
6. Accuracy of Movement or Tracing Board (individual apparatus)
7. Aiming (individual paper)
8. Speed of Movement (group paper)

III. Balancing tests

1. Body Balancing (individual apparatus)
2. Stick Balancing (individual apparatus)

IV. Complex eye-hand coordination tests

1. Link's Machine Operator's (individual apparatus)
2. Card Sorting (individual apparatus)
3. Card Assembly (individual apparatus)
4. Packing Blocks (individual apparatus)

V. Assembly tests involving manipulation and responses to spatial relations

1. Stenquist Assembly (group apparatus)
2. Paper Form Board (group paper)
3. Link's Spatial Relations (individual apparatus)
4. Cube Construction (group apparatus)

VI. Tests of mechanical knowledge

1. Stenquist Picture Tests I and II (group paper)

VII. Miscellaneous tests

1. Slow Movement or Motor Inhibition (individual paper)
2. Digit-Symbol Substitution (group paper)
3. Letter Cancellation (group paper)
4. Number Cancellation (group paper)
5. Rhythm or Perception of Time (group apparatus)

It will be of interest to compare the authors' tentative classification of the tests with the functional classification discovered by the present writer's analysis.

The tests were administered to boys who were enrolled in the seventh and eighth grades of a Minneapolis junior high school and whose curriculum included the shop courses. A number of factors determined the selection of this particular group of subjects. Boys were chosen rather than men because it was felt that in boys individual differences would be determined less by differences in the amount of mechanical training and more by stable abilities which accrue from general sources. Boys from the middle class were selected rather than a more heterogeneous sample. The reason for this selection was that in all probability the boys in the upper classes have had a restricted opportunity to acquire mechanical abilities and the boys from the low, under-privileged class have had restrictions in their development, also. Selection of boys who were enrolled in shop courses was determined by the need for a suitable criterion which could not only be reliably determined but which would also suitably represent the varied operations which are considered to comprise mechanical ability. The battery of 26 tests required 11 hours of testing, and complete data were collected for 217 boys.

The Original Analysis of the Variables of the Preliminary Experiment

The authors of the Minnesota study were interested in the organization of mechanical ability because of its implication for this very practical question: Is it possible for an individual to be distinctly gifted in one line of mechanical work and to possess poor or mediocre capacity in others, i.e., is there a

mechanical ability or group of abilities? In their attempt to answer this question, they intercorrelated all of the variables employed in the preliminary experiment and then subjected these intercorrelations to analytical procedures. Since Spearman's theory of general ability, particularly with respect to intellectual functions, was one of the most arresting theories of the day, the authors concerned themselves principally with the task of finding evidence of a general mechanical ability.

In order to fulfill Spearman's demand that only dissimilar tests be subjected to scrutiny when a general factor is sought, the authors carefully examined their own tests and selected a group of 15 tests which were dissimilar to each other and were not correlated with each other to an exceptional degree. In light of current factor theory, the implications of this selective procedure are obvious. Similar tests are likely to be those highly intercorrelated and if several highly intercorrelated tests are encountered in a group of tests, we are certain to find a cluster or a factor. If such a cluster were found, it would not be possible to account for all of the inter-test correlations in terms of one general factor.

The intercorrelations of the selected 15 variables were arranged in the usual tabular form, and the inter-columnar correlations for this table varied from minus .68 to plus .81. It is obviously impossible to find any hierarchical order in such a table. A similar lack of hierarchical arrangement was encountered when the variables were corrected for attenuation. The results of this procedure combined with the results of certain preliminary procedures forced the authors to the conclusion that their data offered no evidence for a general factor in mechanical ability.

The authors then sought evidence for the existence of group factors. On a trial-and-error basis, attempts were made to construct hierarchies by arbitrarily selecting certain tests for grouping. This procedure was adopted because group factors are general within their own restricted field, that is, variables which define a group factor may be arranged in hierarchical order. The authors assumed that if several well-defined hierarchies were disturbed when tests from any other hierarchy

were included, the evidence would be strongly in favor of the existence of group factors, that is of mechanical abilities rather than a general mechanical ability. The writers reported that they found seven perfect hierarchies of four tests each. Two of these hierarchies were found to be mutually exclusive with respect to their component tests. The other five hierarchies drew upon the same tests. The writers do not indicate what variables contribute to all of the hierarchies. Since such hierarchies can be arbitrarily constructed without one's knowing to what degree they overlap, the analysis of the organization of mechanical ability provided by the Minnesota study leaves much to be desired. From their presentation it may be inferred that if a general mechanical ability factor is present, it is certainly not sufficient to account for all of the inter-test correlations. It may be further inferred that group factors do exist, but the degree of independence which they possess and their exact nature is in no way manifest.

A Factor Analysis of the Variables in the Preliminary Experiment¹

The Minnesota investigators had employed numerous variables which are representative of if not identical with many in use today. Their population was quite large and in their attempts at the analysis of the organization of mechanical ability they determined the intercorrelations among all of the variables. An analysis of the Minnesota data using devices which were not then available is in order in two respects: (1) to round out the classic Minnesota investigation and (2) to shed light on the current problems of measurement of mechanical ability. As a consequence the intercorrelations of the Minnesota variables given in Table 1 were submitted to a centroid analysis; seven factors were extracted, six of which are given in Table 2. The centroid analysis was made for twenty-seven variables. The tests were the original list of twenty-six variables with the exception of the cube construction test, which was not intercorrelated by the Minnesota

¹ This analysis is concerned only with the variables employed in the preliminary experiment. In a forthcoming publication additional analyses of the Minnesota data will be reported.

authors with the other variables. This left twenty-five tests. The *Stenquist Picture Test*, however, really comprised two parts, I and II. Each was scored separately, correlated with the other variables separately, and the two parts are treated as two variables in the present study. Age was also included as a variable and thus the total list of twenty-seven variables which the present writer has analyzed is comprised.

TABLE 2
Centroid Matrix

	A	B	C	D	E	F	h^2
1	.00	.25	-.19	-.37	-.26	-.08	.32
2	.29	.06	.26	.14	.03	-.20	.21
3	.34	-.46	.35	.25	.43	.25	.76
4	.31	.09	.23	.26	-.22	-.09	.28
5	.50	.18	-.24	.31	-.10	.13	.46
6	.53	.12	-.22	.49	-.10	.20	.63
7	.48	.26	-.09	.22	.22	.09	.41
8	.44	.16	-.35	-.10	.48	-.25	.64
9	.50	.27	-.06	.17	-.15	.18	.41
10	.52	-.30	-.28	.08	-.02	.07	.45
11	.40	.28	-.38	-.13	.33	-.37	.64
12	.35	-.46	.31	.18	.29	.27	.62
13	.49	-.05	-.20	.18	.10	.15	.35
14	.48	-.47	-.17	-.04	-.11	.08	.50
15	.14	.06	.15	.20	.10	-.09	.10
16	.23	-.11	-.22	-.05	.10	-.08	.13
17	.44	.29	.10	-.27	.03	.11	.37
18	.27	.08	.23	.08	-.04	-.26	.22
19	.43	-.36	-.25	-.28	-.18	-.04	.48
20	.30	-.42	-.18	-.29	-.20	-.15	.44
21	.57	-.56	-.15	-.22	-.06	.05	.71
22	.20	.29	.15	.13	-.21	-.03	.20
23	.33	.32	.17	-.32	.14	.16	.37
24	.34	-.38	.08	-.44	-.16	.34	.55
25	.44	.29	.17	-.35	-.18	.23	.50
26	.30	-.11	.06	.17	-.27	-.32	.30
27	.19	-.04	.23	.05	-.11	-.27	.17

The centroid factors were orthogonally rotated according to the well-known principles of maximizing the number of zero loadings and minimizing the number of negative ones. After thirty-six sets of rotations had been made it was found that further rotations would contribute little to the simplicity of the patterns revealed. Each of the six orthogonal factors is well defined, and for the first five factors the loadings are exceptionally high. The rotated factor matrix is given in Table 3 and the transformation matrix is given in Table 4.

TABLE 3
Rotated Factor Matrix

	I	II	III	IV	V	VI	h^2
1 Age ..	16	28	-.44	-.04	.02	.19	.34
2 Aiming ..	-.12	13	.07	.17	.11	-.33	.19
3 Army Alpha ..	.04	-.10	.85	.04	.11	-.02	.75
4 Body Balancing ..	-.10	15	.12	.34	-.12	-.32	.28
5 Card Assembling, Time	.05	10	.05	.66	.07	.08	.46
6 Card Sorting, Time	.02	01	.17	.77	-.00	.09	.63
7 Digit-Symbol ..	-.12	17	.18	.48	.31	.10	.41
8 Letter Cancellation ..	.08	04	-.02	.23	.76	.06	.64
9 Link's Machine Operator's	.00	31	.07	.55	.01	.05	.41
10 Link's Spatial Relations, Both Boards	.46	-.05	.22	.40	.15	-.09	.46
11 No Cancellation ..	.06	.08	-.22	.26	.71	-.03	.64
12 Otis Criterion Score ..	.13	-.04	.77	.05	.03	-.02	.62
13 Packing Blocks ..	.18	.03	.24	.45	.19	.11	.35
14 Paper Form Board ..	.59	-.02	.28	.25	.04	-.07	.50
15 Rhythm ..	-.17	.00	.15	.14	.08	.15	.10
16 Slow Movement ..	.05	.12	.24	-.02	.12	-.20	.13
17 Speed of Movement ..	.06	.55	.06	.15	.21	.04	.39
18 Steadiness, No. of Contacts	-.08	15	.07	.14	.11	-.37	.21
19 Stenquist Assembly ..	.66	.10	.04	.13	.11	-.09	.49
20 Stenquist Picture I ..	.63	.12	.00	.00	.07	-.19	.44
21 Stenquist Picture II ..	.74	.06	.34	.15	.14	-.10	.71
22 Stick Balancing ..	-.17	.26	.06	.27	-.10	-.16	.21
23 Tapping A ..	.10	.56	.09	.01	.23	.11	.39
24 Tapping B ..	.10	.73	.06	.07	-.00	.12	.56
25 Tapping C ..	.12	.69	.03	.13	.00	.04	.51
26 Tracing Board, No of Errors	.14	-.01	-.01	.28	-.03	-.46	.31
27 Tracing Paper, No of Errors	.00	.08	.06	.06	.02	-.40	.18

TABLE 4
Transformation Matrix

	A	B	C	D	E	F
I	.35	-.64	-.41	-.45	-.31	.04
II	.45	.49	.39	-.54	-.20	.24
III	.35	-.48	.51	.24	.41	.39
IV	.62	.25	-.35	.61	-.24	.08
V	.35	.15	-.27	-.23	.74	-.42
VI	-.21	.17	-.47	-.11	.29	.78

In order to clarify the implication of these factors, each will be discussed with respect to its component variables and with respect to the degree to which component variables are saturated with other factors.

Since the factors are discussed variable by variable, the data will be presented in the form of factorial equations.²

Factor I—Spatial Visualization

	I	II	III	IV	V	VI	U^2
10. Link's Spatial Relation21	.00	.05	.16	.02	.01	.55
14. Paper Form Board35	.00	.08	.06	.00	.00	.51
19. Stenquist Assembly44	.01	.00	.02	.01	.01	.51
20. Stenquist Picture I40	.00	.00	.00	.01	.04	.55
21. Stenquist Picture II55	.00	.11	.02	.02	.01	.29

The *Spatial Relations Test* of Link is the original form of the currently popular *Minnesota Spatial Relations Test*. The *Paper Form Board Test* from the *Army Alpha* is now modified and known as the *Minnesota Paper Form Board Test*. The *Stenquist Assembly Test* has been lengthened and modified and appears as the *Minnesota Mechanical Assembly Test*. Twenty-one per cent of the total variance of the *Spatial Relations Test* is due to the spatial factor. Another sixteen per cent of its variance is due to factor IV, which is identified as a dexterity factor. Users of the *Minnesota Spatial Relations Test* have long believed that this test calls for spatial ability and for a certain amount of dexterity or manipulative ability as well. The *Paper Form Board Test* has thirty-five per cent of its variance attributable to the Spatial factor; about eight

² Factorial equations are more revealing than simple factor loadings because they not only show how much of the total variance of a test is due to each factor, but they also indicate how much of the variance is not due to common factors, i.e., how much is unique (U^2) to the test in the present test sample; the values in the factorial equations are equal to the respective factor loadings squared.

per cent is due to factor III, which in this study is called the scholastic ability factor. The *Assembly Test* is largely a measure of spatial ability. It is very interesting to observe that it calls for the dexterity factor, IV, to a negligible degree. Contrary to the implications of its name, the *Assembly Test* calls for the ability to visualize the relations of parts rather than the ability to excel in the manual process of assembling parts.

Some explanation should be offered for the difference between the saturations of tests 20 and 21, both of which are Stenquist picture tests and contribute greatly to the spatial factor. It is observed that test 21 owes approximately 10 per cent of its total variance to the scholastic ability factor. A possible explanation of this difference between the two picture assembly tests is that test 21 was given under marked restriction of time. The time for each exercise in the test is reduced twenty per cent. The meaning of time restriction for this saturation of test 21 is a question that will be discussed further under factor III.

Factor II is determined by tests which appear to call for a high degree of speed in certain stereotyped ballistic movements of the wrist and forearm.

Factor II—Stereotyped Movement

	I	II	III	IV	V	VI	U^2
(9 Link's Mach Operator) ..	.00	.10	.01	.30	.00	.00	.59
17. Speed of Movement00	.30	.00	.02	.04	.01	.63
23. Tapping A00	.31	.01	.00	.05	.01	.62
24 Tapping B01	.53	.00	.01	.00	.02	.43
25 Tapping C02	.48	.00	.02	.00	.00	.48

The *Speed of Movement Test* calls for as many vertical marks as possible within a time limit. Variables 23, 24, and 25 were all tapping tests which were conducted in different ways. In test 23, the individual is to make one dot in each of the large number of tiny squares. In test 24 he is to make as many dots as possible in each of the large squares outlined on his test paper, fifteen seconds allowed per square. In test 25, the individual is equipped with a metal plate and a stylus and his rate of tapping is determined. In all four of these tests the individual must make very rapid movements of the

wrist and the forearm. It is observed that test 9 had ten per cent of its total variance attributable to this factor. The machine operator's test, 9, involves dropping balls into a funnel at intervals which permit the ball to drop through a momentarily appearing aperture in a platform beneath the funnel. This performance calls for a vertical dropping movement of the hand which is not unlike the vertical movements called for in the various tapping tests.

Factor III is called the Scholastic Ability Factor for the simple reason that two well-known tests of scholastic ability comprise it.

Factor III—Scholastic Ability

	I	II	III	IV	V	VI	U^2
(1. Age)03	.08	-.20	.00	.00	.04	.65
3 Army Alpha00	.01	.73	.00	.01	.00	.25
12. Otis02	.00	.59	.00	.00	.00	.39
(21. Stenquist Picture II)55	.00	.11	.02	.02	.01	.29

These tests are the *Army Alpha* and the *Otis*. The scores are not intelligence ratios but are simply raw scores. As would be expected, this factor is negatively correlated with age. It will be remembered that the subjects for the preliminary experiment were picked from the seventh- and eighth-grade classes. Ordinarily the older children in a public school class are the children who have been retained, and most frequently this retention has been due to dullness on the part of the child. If this general observation may be validly applied here it accounts for the negative relation between this factor and age. It is further observed that variable 21, the *Stenquist Picture Test II*, has a probably significant relationship with this factor. The next variable most highly correlated with this factor is variable 14, the *Paper Form Board Test*. The *Army Alpha*, the *Otis*, the *Paper Form Board*, and the *Stenquist Picture Test II* are all pencil-and-paper tests which may be classified as mental tests and all four are highly speeded. Perhaps these common characteristics can account in part for the positive intercorrelation.

Factor IV is determined by tests which appear to call for a type of manual dexterity.

Factor IV—Manual Dexterity

	I	II	III	IV	V	VI	U^2
(4. Body Balancing)01	.02	.02	.11	.02	.01	.72
5. Card Assembly, Time00	.01	.00	.43	.01	.01	.54
6. Card Sorting, Time00	.00	.03	.60	.00	.01	.36
7. Digit-Symbol02	.03	.03	.23	.10	.01	.58
9. Link's Mach Operator00	.10	.01	.30	.00	.00	.59
10. Link's Spatial Relation21	.00	.05	.16	.02	.01	.55
13. Packing Blocks03	.00	.06	.21	.04	.01	.65

Variables 5 and 6 are the best measures of this factor and they are the well-known Card Sorting and Card Assembly tests. The next best test appears to be test 9, which is the *Machine Operator's Test* described in the discussion of Factor II. Tests which have correlations of approximately .30 or greater with this factor tend to contribute to its meaningfulness; variables 4 and 7 are possible exceptions. There is no apparent reason why variable 4 should be expected to correlate with the factor of manual dexterity, and it is somewhat surprising although not wholly unreasonable to find a test of digit-symbol-substitution highly correlated with the dexterity factor. The exact nature of the manual dexterity factor is not perfectly clear. To what degree excellence of performance on tests of this factor is due to a type of acuity of perception and recognition has not been determined. This question will be discussed further with presentation of Factor V. Test 10, the *Spatial Relations Test*, which owes sixteen per cent of its variance to this factor, obviously requires manipulative skill and would be expected to draw upon some sort of manual dexterity to a significant degree.

Factor V is identified by tests which have been termed perceptual tests and have been found by previous investigation to determine a perceptual factor (9).

Factor V—Perceptual Factor

	I	II	III	IV	V	VI	U^2
(7. Digit-Symbol)02	.03	.03	.23	.10	.01	.58
8. Letter Cancellation01	.00	.00	.06	.57	.00	.36
11. Number Cancellation00	.01	.05	.07	.51	.00	.36

The two cancellation tests, no. 8 and 11, have remarkably high saturations with this factor. Test no. 7, the *Digit-Symbol Substitution Test*, has unexpectedly low saturation. The reason for this test's having a high saturation with the

dexterity factor rather than with the perceptual factor may actually be due to an undetected clerical or computational error. On the other hand, it is conceivable that the relations observed here are valid relations and perhaps some unrecognized aspect of the digit-symbol performance could, if understood, give insight into the nature of manual dexterity. It should be observed in passing that tests commonly employed in the measurement of perceptual ability, including cancellation tests such as appear in this battery, call for speed of simple, routine perception and the emphasis is not upon acuity of differential recognition and discrimination. Perhaps manual dexterity is a complex ability which does call for a differential recognition and discrimination which is not purely or primarily of a routine nature. The hypothesis that tests for manual dexterity draw upon two distinct abilities, a manipulative ability and a recognition ability, may be tested by determining the factorial composition of manipulative tests when given in a darkened room or under conditions to eliminate visual discrimination and comparing the results with the factorial composition of these tests when they are administered under ordinary conditions. Factor analysis is not necessary to test this hypothesis, however, and other, simpler designs would undoubtedly be more economical.

Factor VI has low saturations. The relatively low inter-correlations among the tests which contribute to the steadiness factor are in part due to their unreliability. The tests with the highest loadings have certain common characteristics, and as a consequence the factor of steadiness has been hypothesized.

Factor VI—Steadiness

	I	II	III	IV	V	VI	U^2
2. Aiming01	.02	.01	.03	.01	.11	.81
4. Body Balancing01	.02	.01	.11	.01	.10	.74
18. Steadiness, No. contacts01	.02	.01	.02	.01	.14	.79
26. Tracing Board, No. errors . .	.02	.00	.00	.08	.00	.21	.69
27. Tracing Paper, No. errors . .	.00	.01	.00	.00	.00	.16	.83

Body Balancing calls for the individual's balancing himself on the ball of one foot while standing on a three-inch cube of wood. Buxton (2) reported a factor of manual steadiness based upon two tests, one of which was a thrusting test which

calls for an operation similar to the nine holes steadiness test. In a later paper, Seashore *et al.* (7) reported a steadiness factor which was based upon measurements of postural sway. The steadiness factor which is reported in this study is dependent upon a measure of postural steadiness, body balancing, as well as measures of manual steadiness. These data in combination with the data provided by other observers (4, 6, 8) suggest that there is one general steadiness factor (probably due to basically physiological individual differences conceivably in proprioceptive sensitivity or muscular tonus). The possibility that there exists only one general factor of steadiness has quite useful implications and the hypothesis may readily be tested.

The conclusion is apparent from this analysis that no common factor could account for the intercorrelations among the various tests. On the contrary it is apparent that the inter-test correlations may be accounted for in terms of six independent and meaningful factors. Obviously such factors cannot be considered immutable or part of the universe, but among tests which the authors of the Minnesota study (as well as current investigators) consider to be important and representative, meaningful, well defined, independent, functional groupings have been demonstrated.

Other Analyses of Mechanical Ability

There are few factorial studies relating to mechanical ability reported in the literature; the most satisfactory one is Harrell's, which he succinctly summarizes as follows: "The intercorrelations of thirty-seven variables, including the Minnesota battery of 'mechanical ability' tests, the seven MacQuarrie tests of 'mechanical ability,' O'Connor's Wiggly Blocks, and the Stenquist picture-matching test, were analyzed by Thurstone's centroid method. Five factors, Perceptual, Verbal, Youth, Manual Agility, and Spatial, were taken out. Factors prominent in so-called mechanical ability tests are the Spatial and Perceptual ones with MacQuarrie's dotting test significantly high in the Manual Agility factor. Each of the factors can be measured with group pencil-and-paper tests"

(3). Harrell's Factor IV, agility, could readily be identified with the present writer's manual dexterity factor, IV, were it not for the participation of the dotting test in Harrell's agility factor. Since there were no additional dotting tests in his battery or other tests of the ballistic type, we cannot be confident that the loading of the agility factor with this particular test has general significance. It may be suspected that this test is highly related with a "ballistic" or stereotyped movement factor; high correlations between MacQuarrie's dotting and tapping tests have been reported (1).

In "*The Application of Multiple Factorial Methods to the Study of Motor Abilities*" (2) Buxton found several factors which are in essential agreement with the results of the present study. His Factor I is a dexterity factor which appears to be identical with the dexterity factor described in the present study. His Factor III, a steadiness factor, is determined by manual tests only. Buxton's Factor VI is found in neither the present study nor in Harrell's study. It appears to call for general coordination of the muscles of the forearm, upper arm, and shoulder girdle.

A "*Multiple Factorial Analysis of Fine Motor Skills*" by Seashore, Buxton, and McCollom has been reported (7). The results of this study have uncertain value for the broader problem of mechanical ability for several reasons. The tests employed are of a rather specific nature and may not in general be considered representative of tests for mechanical ability. The analysis was based upon a sample of 50 men, a small sample for factor analysis. Moreover, it appears likely that the authors might have extracted additional factors; from an examination of the factor matrix it does not appear that the zero-order correlations may be satisfactorily reproduced. As the writers themselves note, some of the factor loadings are inconsistent with the identifications of the tests and factors. Although the study was not intended to contribute directly to the problem of the nature of mechanical ability, it does support the implications of the present study by suggesting the existence of a postural steadiness factor and a factor of stereotyped movements of the arm such as are called for by Factor III in the present study.

The literature offers several interesting studies of the organization of strength, agility, and motor fitness. In a comprehensive approach to the study of the nature of mechanical ability, cognizance must be given to such possible components. This aspect of mechanical ability will be treated in a forthcoming study.

Discussion

The occurrence of an orthogonal factor structure is of some interest because both Buxton and Harrell have found an oblique structure to be most appropriate in analyzing their variables. The orthogonal feature of the present analysis may be in part due to the age of the subjects. They are seventh- and eighth-grade boys; the subjects for the other analysis of Buxton and Harrell had been adults. Babcock and Emerson, in a study of MacQuarrie's test for mechanical ability, have found intercorrelations among the subtests to increase in magnitude with age of subject (1). This interesting tendency is quite the reverse of that usually observed for the so-called mental tests. The tendency for mental abilities as defined by factor analysis to become increasingly independent of each other with increasing age and the apparent tendency of mechanical abilities to become more highly related with each other with increasing age requires explanation, and it is possible that the final answer to this paradox may yield some insight into the nature of the respective classes of ability. Since all the mechanical tests employed in this study (and in other studies too) are not equally valid, it cannot be argued that excellence in practical mechanical work requires a high degree of all of these abilities and that therefore a type of automatic selection occurs which eventually results in those who have a high degree of all abilities being active in mechanical operations and those deficient in any one being equally discouraged in all. It is not likely that any selective situation such as this can be invoked to explain this paradoxical trend. A slightly modified hypothesis may be offered, however; namely, that a high order of spatial ability is necessary for personally gratifying performance in any mechanical operation and that persons who are deficient in spatial ability may in general tend to

avoid mechanical work or participate with little zeal and as a consequence neglect the development of any purely manual facilities with which they might have started life. The apparent anachronism in the development of mechanical ability is provocative of speculation. Although the trend may not be verifiable its implications are important enough to warrant scrutiny of any suitable genetic records which may exist.

It is of interest to compare the classification of the tests arbitrarily imposed by the authors at the beginning of the Minnesota investigation and the classification revealed by the empirical evidence offered by the intercorrelations and by subsequent factor analysis. The first arbitrary classification was Intelligence and this classification is validated by the results of the statistical analysis. The second classification, Simple Motor Tests, actually includes measures for two factors defined by the present factor analysis, i.e., Factors I and III, stereotyped movement and steadiness. The third group comprises the two balancing tests and is also included in the steadiness factor. The fourth group, Complex Eye-Hand Coordination Tests, comprises tests which determine Factor IV, identified in the present study as manual dexterity. The participation of the *Digit-Symbol Substitution Test* and of other measures in this factor suggests that the designation eye-hand coordination for this class of ability may be most appropriate. As previously mentioned, however, further research is required before the exact nature of manual dexterity can be determined. It may well be that manual dexterity is an operational unit which when analyzed for logical content does demand a type of manipulative ability and visual discrimination. It is equally possible, however, that manual dexterity or eye-hand coordination may be fractionated into two relatively independent components, visual discrimination and a manipulative ability. Group V, designated as Assembly Tests Involving Manipulation and Response to Spatial Relations, appears to be appropriately identified insofar as our analysis has revealed that these tests, though primarily dependent upon the spatial visualization ability, draw in varying degrees upon manual dexterity. This class probably was regarded as a

combination group, however; i.e., it is not apparent that the Minnesota authors foresaw that such tests as the *Spatial Relations Test* and the *Assembly Test* would be primarily determined by the same factors as the *Paper Form Board Test*. Demonstration that spatial ability is a primary component of such manual tests as the *Stenquist Assembly* as well as the *Paper Form Board Test* was probably first made by Harrell. The *Stenquist Picture Tests*, set aside in a separate class as requiring mechanical knowledge, are found actually to contribute to the spatial factor. Apparently the perceptual factor was entirely unanticipated by the Minnesota authors; the cancellation tests are classified under miscellaneous.

Summary

The present study comprises a review of the preliminary Minnesota experiment and a factorial analysis of the variables. The analysis reveals that interrelationships of the variables are of such a nature as to yield upon analysis six meaningful independent factors. For the most part the loadings are exceptionally high and the results of the analysis are exceptionally definitive. The factors are:

- I—Spatial Visualization
- II—Stereotyped Movement
- III—Scholastic Ability
- IV—Manual Dexterity
- V—Perceptual—Speed
- VI—Steadiness

REFERENCES

1. Babcock, H. and Emerson, M. R. "An Analytical Study of the MacQuarrie Test for Mechanical Ability." *Journal of Educational Psychology*, XXIX (1938), 50-55.
2. Buxton, C. "The Application of Multiple Factorial Methods to the Study of Motor Abilities." *Psychometrika*, III (1938), 85-93.
3. Harrell, W. "A Factor Analysis of Mechanical Ability Tests." *Psychometrika*, V (1940), 17-33.
4. Humphreys, L. G., Buxton, C. E., and Taylor, H. R. "Steadiness and Rifle Marksmanship." *Journal of Applied Psychology*, XX (1936), 680-688.
5. Patterson, D. G., Elliott, R. M., Anderson, L. D., Toops, H. A., Heidbreder, E. *Minnesota Mechanical Ability Tests*. The University of Minnesota Press, Minneapolis, 1930.

6. Seashore, R. H. and Adams, R. D. "The Measurement, Steadiness. A New Apparatus and Results on Marksmanship." *Science*, LXXVIII (1933), 285.
7. Seashore, R. H., Buxton, C. E., and McCollom, I. N. "Multiple Factorial Analysis of Fine Motor Skills" *American Journal of Psychology*, LIII (1940), 251-259.
8. Spaeth, R. A. and Dunham, G. C. "The Correlation Between Motor Control and Rifle Shooting." *American Journal of Psychology*, LVI (1921), 249-256.
9. Wittenborn, J. R. "Factorial Equations for Tests of Attention." *Psychometrika*, VIII (1943), 19-35.

SUGGESTIONS FOR THE CONSTRUCTION OF MULTIPLE-CHOICE TEST ITEMS

CHARLES I. MOSIER, M. CLAIRE MYERS, AND HELEN G. PRICE
State Technical Advisory Service, Social Security Board, Washington, D. C.

IN WRITING items in a particular area it is possible, but not very profitable, to use the inspiration technique. One reads until he is inspired to write an item, jots it down, and then reads some more. Those concepts that fall readily into item form get tested over and over again; those more difficult to test go untested. This procedure is likely to result, among other things, in very spotty coverage of the subject-matter area. In writing items, as in other activities, planning is essential.

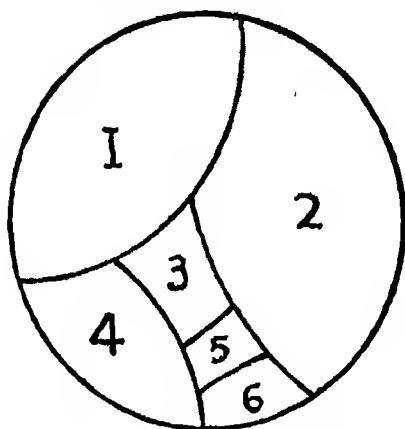
The present paper presents three work tools for the construction of multiple-choice items: definition of the subject-matter area to be covered and systematic sampling of it; a check list of the kinds of questions which can be asked; and a summary of criteria for multiple-choice items.

Although these materials were developed for use in examining for public personnel selection, they are applicable, with minor changes, in the other fields of test construction.

To cover a subject-matter area adequately, whether in educational testing, merit system examining, or aptitude testing, the first step should be a definition of the area or areas to be tested. Even in a limited field, all possible questions cannot be asked; it is necessary to resort to a sampling of the field. If, however, the individual's performance on the sample is to represent his performance in the entire field, it is a truism that the sample must represent the field; the sample to be representative cannot be left to chance, but should be planned. The definition of the field states the limits which must be reached; it does not prohibit going beyond those limits. If items are written outside the boundary of the definition, they

can usually still be used; if no items are written on an area within the field, that part of the area has not been sampled. A picture may help to visualize this.

The entire circle represents the particular subject-matter area, e.g., knowledge of elementary and intermediate statistics. Each subdivision represents a set of related concepts within the broader field. For example, 1 might be the concepts relating to frequency distributions; 2 those relating to central tendency; 3 those relating to dispersion, etc. Unless the total area is defined by correctly drawing the total circle, e.g., closing the circle in the lower right quadrant, the existence of the sixth subdivision escapes our attention.



It is essential, if one is to test adequately for a knowledge of the principles and techniques in statistics, that he be prepared to include test items in each of the areas. It is not necessary to include one from every area in every examination; one should not, however, include three from section 1 and none from section 6 unless he is prepared to say that the concepts in 1 are very important and those in 6 are inappropriate to the purposes of the particular examination.

After the area has been outlined, the next step is the identification of its significant subdivisions. This can be accomplished by listing the important concepts, topics, principles, or subdivisions of the field or of the various types of skills which contribute to the total. This list represents the specifications

for the set of items. It should not, generally, be a list of the items to be written, but of the sub-areas to be covered, with each area to be sampled by several items. Preparing such a list assumes either an authoritative outline already set up, or an over-all familiarity with the field. The chapter headings in a text or the paragraph headings within a chapter constitute one source of such a listing. Once made the list is important: first, as an index to the resulting assembly of items; second, as a guide to future item construction; and third, as a more elaborate definition of the types of items contained in the subject-matter area. Table 1, showing an outline of statistics, is presented as an illustration, not necessarily as a model.

TABLE 1
Outline of Statistics

1 General	5 Dispersion
Interpretations	General
Life Tables	Sigma
"Severity Rate"	Average deviation
Simple computations	Quartile deviation
Sources and bibliographies	6 Correlation
Tabular presentation of results	Interpretation and use
Terms and symbols	Scatter diagram
Planning surveys	Pearson
Compound interest formula	Other coefficients
2 Frequency Distributions	Regression lines
Class intervals and limits	Partial and multiple
Bimodality	Coefficient of alienation
Symmetry	Standard error of estimate
Skewness	7 Non-Linear Regression
Frequency polygon	Trend lines
Kurtosis	Population curves
Percentiles	Eta and Blakeman's test
Normal curve properties	8 Sampling
Binomial expansion	Methods: random, weighted
3. Charts, Graphs and Index Numbers	Sampling errors
Ogive	Interpretation, S E, and C R.
Moving average	Combination of samples
Straight-line equation	Probability theory
Semi-log paper	
Interpolation and extrapolation	
Pictograms and circle diagrams	
Gantt, Lorenz, ratio, and time charts	
Graphic computation	
Index numbers	
4. Central Tendency	
Mean	
Median	
Mode	
Harmonic Mean	
Geometric Mean	

Once the list of concepts, principles, skills or topics to be tested has been completed, appropriate sources or sections of the source should be used in constructing the items.

It is useful to distinguish between the concept, or the skill, being tested and the kind of question asked to test that concept. A further distinction which should be made is between the kind of question asked (e.g., what, why, who?) and the form of the test item used to ask the question (e.g., true-false, multiple-choice, completion). We are concerned here primarily with multiple-choice items and the *form* of the test item will not be considered further. In Table 2 is a check list of some of the kinds of questions which may be asked together with illustrative examples based on the concept of central tendency.

TABLE 2
Types of Questions

1 Definition	
a	What means the same as .
b	What conclusion can be drawn from .
c	Which of the following statements expresses this concept in different terms? Example: The value which is determined by adding all of the scores and dividing by the number of cases is known in statistics as the (1) arithmetic mean; (2) median, (3) mode, (4) harmonic mean, (5) average deviation
2. Purpose	
a	What purpose is served by .
b	What principle is exemplified by
c	Why is this done
d	What is the most important reason for Example The mean is obtained for the purpose of providing: (1) a single number to represent a whole series of numbers, (2) the central point in a series, (3) a measure of group variability; (4) an indication of the most frequent response given, (5) an estimate of the relationship between two variables
3. Cause	
a	What is the cause of
b	Under which of the following conditions is this true Example: From which of the following measures of central tendency will the sum of the deviations equal zero? (1) the mean; (2) the mode, (3) the median; (4) an arbitrary origin, (5) any measure of central tendency

4 Effect

- a. What is the effect of .
- b. If this is done, what will happen?
- c. Which of the following should be done (to achieve a given purpose)?
 Example: The arithmetic mean of 55 cases is 83.0. If 3 of the cases, with values of 82, 115, and 130 are deleted from the data, the mean of the remaining 52 cases will be (1) 81.50; (2) 77.05, (3) 83.00, (4) 84.50; (5) 94.08

5 Association

What tends to occur in connection (temporal, causal or concomitant association) with .

Example. If the distribution of scores is skewed positively, the mean will be

- (1) lower than the median,
- (2) the same as the median,
- (3) higher than the median,
- (4) relatively unaffected,
- (5) the same as the mode

6 Recognition of Error

Which of the following constitutes an error (with respect to a given situation)?

Example: The mean should not be used as the measure of central tendency when.

- (1) the distribution of scores is significantly skewed,
- (2) there are a large number of cases;
- (3) a non-technical report is to be prepared,
- (4) the data are continuous,
- (5) other statistical formulas are to be computed

7 Identification of Error

- a. What kind of error is this?
- b. What is the name of this error?
- c. What recognized principle is violated?

Example In computing the mean of a distribution from grouped data, the sums of the deviations above and below the arbitrary origin were found to be 127 and 189, respectively. The final value for the mean was in error. Of the following possibilities, that one which is most likely to have caused the error is that the computer:

- (1) failed to note the correct sign in adding the mean of the deviations to the assumed origin;
- (2) used an assumed mean higher than the true mean;
- (3) omitted some of the cases in tabulating the data;
- (4) divided by the wrong number of cases,
- (5) multiplied by the wrong class interval value

8. Evaluation

What is the best evaluation of . . . (for a given purpose) and for what reason?

Example: When the number of cases is small, e.g., less than 20, and the magnitude of the values is likewise small, the use of an assumed mean in the computation of the mean can best be evaluated as.

- (1) less efficient than computation from actual values,
- (2) likely to distort the value obtained by the introduction of a constant error,
- (3) more accurate than the use of actual values,
- (4) neither better nor worse than computation by other methods,
- (5) applicable only if the distribution is reasonably symmetrical.

9. Difference

What is the important difference between . . .

Example: Of the following statements, that one which characterizes the essential difference between the mean and the median as measures of the central tendency of a distribution is that

- (1) the magnitude of each score does not contribute proportionately to the computation of the median but does for the mean,
- (2) the median is a point whereas the mean is a distance,
- (3) the mean is less affected by extreme values,
- (4) the median is easier to compute;
- (5) the median is more generally used

10. Similarity

What is the important similarity between .

Example The mean and median are the same in that they are both measures of:

- (1) central tendency,
- (2) distance,
- (3) position,
- (4) variation,
- (5) relationship

11 Arrangement

In the proper order, (to achieve a given purpose or to follow a given rule) which of the following comes first (or last or follows a given item)

Example. In computing the mean for data already grouped in class intervals the most efficient first step is to.

- (1) determine the arbitrary origin and enter the deviation values;
- (2) find the midpoints of the class intervals,
- (3) multiply the frequency in each interval by the midpoint of the interval,
- (4) add the column of scores;
- (5) find the reciprocal of the total number of cases

12. Incomplete Arrangement

In the proper order, which of the following should be inserted here to complete the series?

Example: In deriving the formula for computing the mean from grouped data using an arbitrary origin the following steps were taken.

- a. $X' = \frac{X - A}{i}$,
- b. $\Sigma X = i \Sigma X' + NA$,
- c. $\frac{\Sigma X}{N} = A + \frac{i \Sigma X'}{N}$.

The step which is implied between steps (a) and (b) is:

- (1) solving (a) for X ,
- (2) summing (a) over the N cases;
- (3) multiplying by i ,
- (4) adding A to both terms of (a),
- (5) dividing by N .

13. Common Principle

All of the following items except one are related by a common principle:

- a. What is the principle?
- b. Which item does not belong?
- c. Which of the following items should be substituted?

Example: All except one of the following items (arithmetic mean, median, mode, and quartile) are measures of central tendency; of the following statistics, that one which could be substituted in the series for the item improperly included is the:

- (1) harmonic mean for quartile,
- (2) average deviation for mode;
- (3) range for quartile;

- (4) standard deviation for quartile;
- (5) 50th percentile for median

14 Controversial Subjects

Although not every one agrees on the desirability of ———, those who support its desirability do so primarily for the reason that.

Example. Although not every one agrees that the mean is the best measure of central tendency, those who advocate its general use base their recommendation on the fact that the mean

- (1) has the smallest sampling error,
 - (2) is easiest to compute;
 - (3) is most readily understood,
 - (4) is the most typical score,
 - (5) is not affected by extreme values
-

Frequently a concept can be tested by a variety of questions; for some concepts only one kind of question is appropriate; for still others, several are applicable, but one or two are clearly most appropriate. To secure adequate sampling in the construction of items on any concept, the concept should be checked against the list and as many items written as seem appropriate, each asking a different kind of question. By writing items which test knowledge of a concept through several kinds of questions, it is frequently possible to sample the area at several levels of difficulty, ranging from the simplest to the most difficult.

The construction of a number of different items on each concept included in the outline of the subject-matter field provides a more effective means of meeting the objective of those who feel that internal weighting is essential to give the appropriate emphasis to the more important concepts. The proponents of internal weighting argue that since certain concepts are more important, extra credit should be given for the questions on those concepts. If the desirability of such extra credit is granted, the weighting is more easily and more reliably accomplished by including extra questions than by doubling the credits for a single question, thereby doubling the effect of measurement errors in the question. The availability of several different questions on each concept makes it possible to include more than one for any that are considered more important. These considerations apply whether items are being constructed for a single examination or for a central file.

The check list, *Types of Questions*, in Table 2 is not in-

TABLE 3

*Criteria for Constructing Multiple-Choice Items¹***I. General Validity**

- 1 Can we readily predict that those candidates who know the answer would, on the average, be better qualified for the purpose at hand than those who do not?
- 2 Is the item thought-provoking, rather than calling merely for scraps of information?

II Item Content

- 1 Is the content of the item important enough to justify a question—not so specialized that only a few highly selected experts could know the answer?
- 2 Does the subject matter of the item appear to be reasonably related to some type of activity appropriately covered by the test?
- 3 Are the subject matter and the phrasing of the item such that no emotional antagonism will be aroused on the part of the public and the candidate? How would the item look in a newspaper unfriendly to the merit system? (Although this criterion is primarily formulated for merit system examiners, it has application in education testing, there is no need to antagonize students unduly in testing them.)
- 4 Is the content of the item something which can be learned without having actually been on the job itself?
- 5 Does the subject matter mirror a consensus of current authoritative beliefs and opinions, rather than the opinion of one individual?
- 6 Could general terms be used to make the item usable in more than one testing situation. If the principle is applicable only in one situation, has this fact been flagged by specific reference in the item?
- 7 Does the item call for a knowledge of concepts, reasons, and relationships, rather than for mere factual information, whenever the former is appropriate?
- 8 Could any part of the item, such as modifying phrases, qualifications, etc., be omitted without significantly influencing the distribution of responses?

III Item Structure**A. Premise**

- 1 Is a definite task clearly and unambiguously set, so that all candidates work on essentially the same problem? Is it sufficiently clear that an informed candidate could give the correct answer from the premise if it were written as a completion item with no choices given?
- 2 Is the idea stated clearly and directly, with the answer an important part of the statement—not buried at the end of a preposition in a parenthetical “which clause”?
- 3 Does the premise, combined with each choice, constitute a complete unit of thought, both ideologically and grammatically?
- 4 Does the premise, combined with the right answer, constitute a definite and true concept? Is the premise phrased to require something more (the correct choice) for completeness?
- 5 If the question is stated negatively, can it be restated in positive terms? If not, has the attention of the candidate been directed toward the negative phrasing of the question?
- 6 Is the premise phrased to ask for the best rather than the correct answer whenever possible?
- 7 Is the premise stated so complexly that the item becomes a test of whatever is taken to understand a complex premise rather than a test of knowledge or reasoning? If so, is this what was intended?

¹ These criteria, stated in terms of merit system examining, can be made applicable to other testing situations by appropriate changes in wording

TABLE 3

Criteria for Constructing Multiple-Choice Items

-
8. Does the premise avoid all unnecessary content which might give away the answers to other items?
 9. If there are other answers conceivably as good as the intended answer, is the premise limited by the phrase, "of the following"?
- B. Correct Answer
1. Is there one and only one correct answer to the problem as set by the premise? If the intended answer is merely the best of those given, is this indicated clearly in the premise?
 2. Do competent authorities agree on the correct answer?
 3. Could the candidate distinguish the correct answer from the incorrect answers without having read the premise? If so, the item is in need of revision
 4. Does selecting the correct choice require a real or reasoned understanding of the concept rather than mere recall or recognition?
 5. Is there no possibility that a candidate may select the correct response simply because it is the only one containing the same words or phrases as the premise or because of other external characteristics?
- C. Distracters
1. Are the distracters such that a person likely to be inferior on the job will think they are correct? Is each distracter so plausible that someone not knowing the correct answer will choose it?
 2. If there are popular misconceptions in the field, are the distracters designed to attract candidates holding those misconceptions? Can familiar or stereotyped phrasing in the distracters be used to make them sound plausible? Can words or phrases similar to those in the premise be deliberately planted in the distracters to give them added plausibility? Can this increased plausibility be achieved without distracting too many of the qualified candidates as well? Are specific determiners, such as "always" or "never," avoided?
 3. Are the distracters, as well as the intended answer, related to the premise, both ideologically and grammatically?
 4. Are there no distracters so nearly correct that well qualified persons are likely to accept them and be able to defend their answers?
 5. Can the question be answered by someone who knows, not that the intended answer is correct, but that the incorrect choices are obviously wrong? If so, does the item remain at the difficulty level originally intended?
 6. Do all the choices constitute possible answers to a single direct question implied in the premise?
 7. Are the distracters of about the same length and complexity as the correct answer? If not, are there at least two which are parallel with the intended answer?
 8. Do pronouns refer clearly to one and only one antecedent?
 9. Are the choices parallel in grammatical form and in meaning?
 10. Are there repetitions in the choices which could be avoided by putting the repeated thought into the premise?
-

tended to be all-inclusive nor is it intended to prescribe the language to be used in asking any particular kind of question. Rather it is a guide to assist in formulating questions and a check to help insure that all of the appropriate kinds of questions will be asked.

Once the concept to be tested and the kind of questions to be asked have been determined, there remains the task of framing objective questions which will accomplish the intended purpose.

The task can be stated simply: To phrase a question in such terms that (a) all prospective examinees understand the task set; (b) those who have the requisite degree of knowledge will give the intended answer; and (c) all who do not will give another answer. Fashioning such an item is more difficult than formulating the problem. The criteria in Table 3 are set up as aids in constructing such items. They are by no means original; they have been assembled from various sources and are summarized here for convenience. The criteria in the table were formulated for use in merit system examining; each one, however, with slight changes in wording or in emphasis is applicable in other examining settings. These criteria should be applied to each item written; the item should not be considered as complete until the writer is satisfied that it meets each criterion.

After the concept has been identified, the question asked in a premise which is clear and the intended choice written, the next task is that of writing distracters. The ideal distracters are those wrong answers actually given by representatives of the group for which the item is intended (not by some other group), in response to the question asked as a completion item. Since item writers do not normally have ready access to such answers, it is up to them to project themselves into the situation and say, "If I were one of the group to whom this test is to be given and were asked this question as a completion item and didn't know the answer, what response would I give?" On the success with which the item writer can thus project himself into the situation of such candidates and predict their responses rests the value of the item from a technical standpoint.

In presenting a completed item the item writer is, in effect, saying: "This item is to be included in a test and given to a group of examinees, some qualified and some not qualified with respect to a particular objective. If we could accurately separate the qualified individuals from the unqualified, and

examine the responses of each group, we would find the keyed answer given by a relatively high proportion of the qualified group and by a low proportion of the unqualified. Each of the four distracters, on the other hand, should be given by a high proportion of the unqualified group and by a low proportion of the qualified." If, after finishing an item, one cannot in all honesty make this prediction, the item should be worked on further. If this prediction can be made, it still needs empirical verification by observing the actual responses of the two groups, "qualified" and "unqualified," as defined by an acceptable criterion.

•

INFLUENCE OF LEVEL OF MOTIVATION ON THE VALIDITY OF INTELLIGENCE TESTS

L. D. HARTSON

Oberlin College

A TOTAL of 579 students entered Oberlin College during the period 1941-43 for whom IQ's were available from tests taken previous to entering college. These data have been added to those for the group of 734 entering during the years 1934-40, concerning whom a report has previously been made (2), making a total of 1,313 students. These data have been studied for the purpose of checking the earlier results, and examining the relationship between aptitude and scholastic achievement at different levels of motivation.

Follow-Up of Study Reported in 1941

From these 1,313 students a total of 1,444 records are available, since there are 131 cases for whom two IQ figures were reported. For some purposes these might have been averaged, but as that would have prevented their use for comparing the validity of the different varieties of test, each IQ has been treated as though it represented a different student. The same procedure being followed with those who became seniors, the records for 522 students yielded 577 IQ's. The two populations employed in the study are, therefore, 1,444 and 577. The figures used in the computation of high-school scholarship represent not the actual grades but "credit points" obtained by a system for equating different grading systems. College scholarship is handled in centile scores, that for seniors representing the cumulative average for seven semesters.

Correlations were computed (1) between IQ's obtained from the following group tests: *Otis*, *Terman*, *Hennon-Nelson*, *National*, *Kuhlmann-Anderson*, from the individually adminis-

TABLE 1
Correlations between IQ's Derived from Various Tests and (1) High-School Scholarship, (2) 1st-Semester College Scholarship, and (3) OSU Test Scores

Test	N	Scholarship		Fresh. week OSU Test	Test scores		Scholarship			
		High- school	1st-sem college		Mean	σ	High-school		1st-sem college	
										Mean
<i>Obs</i>	680	.328	.360	.513	120.10	8.50	76.49	17.94	48.71	27.91
<i>Terman</i>	304	.285	.390	.542	122.64	10.62	76.34	16.74	49.64	27.73
<i>Hennon-Nelson</i>	258	.385	.404	.536	123.83	9.19	81.68	14.69	50.62	25.44
<i>National</i>	48	.339	.301	.478	129.25	12.52	82.83	12.85	53.42	23.62
<i>Kuhlman-Anderson</i>	63	.250	.165	.396	124.75	11.69	76.56	18.19	55.02	29.73
<i>Stanford-Binet</i>	91	.259	.343	.566	123.66	12.90	71.80	20.14	48.46	30.69
<i>OSU (Pre-Entrance)</i>	331	.349	.511	.774	55.35	26.68	69.02	22.57	41.78	30.45

tered *Stanford-Binet*, and from the *Ohio State University Psychological Examination*, as one variable, and high-school scholarship, as one criterion; (2) between all of these tests and first-semester college scholarship, as a second criterion; (3) between the IQ's on each of these test groups and the *OSU Test* scores. A comparison was also made (4) of the validities of the *OSU Test* scores and those obtained with each of the other tests, with scholarship used as the criterion at both the secondary school and college levels.

The Prediction of High-School and College Scholarship

The results obtained with 1,444 students in general confirm those reported for the smaller group, as may be seen by comparing Tables 1 and 2 with the corresponding tables in the 1941 study. The group tested with the *Henmon-Nelson Examination* again shows a higher validity than do the other tests yielding IQ's (Table 1), even though this group yields a lower validity coefficient than do the students tested with the *Kuhlmann-Anderson* and with the *Stanford-Binet*, when their intelligence is measured by the *OSU Examination* (Table 2).

TABLE 2
Correlations between Various Tests and Scholarship

Test group	N	Scholarship		OSU Test	
		High-school	1st-sem college	Mean	σ
<i>Ohs</i>	680	385	.556	49.24	28.55
<i>Terman</i>	304	319	.564	50.60	28.12
<i>Henmon-Nelson</i>	258	477	.569	48.68	25.32
<i>National</i>	48	399	.621	55.08	26.21
<i>Kuhlmann-Anderson</i>	63	.584	.610	51.69	28.37
<i>Stanford-Binet</i>	91	.482	.584	49.02	28.41
<i>OSU Test (Pre-Entrance)</i>	331	.439	.640	49.46	29.14

This fact offers some suggestion of superiority for the *Henmon-Nelson Test*, although this presumption is not supported by sufficient numbers to make it statistically substantial.

The Mean IQ of the Oberlin Group

Increase in the size of the population studied from 835 to

1,444 represents a change of one point in the mean IQ, or an increase from 121 to 122. The range is 90 to 169, 99 per cent exceeding the general mean of 100. Sigma is 10.7. Comparison with the *Terman-Merrill* sigma, 16.3 (3, p. 37), shows that variability among the Oberlin student-body is much more restricted than it is in the general population. The average IQ for the group who became seniors is the same as for the entire group of freshmen.

Mean College Scholarship of Students at Different IQ Levels

As the number of students who by 1943 had reached senior standing was more than twice the number for whom data were available three years earlier, 577 as compared to 253, a report is again made of first- and seven-semester scholarship of students at different IQ levels (Table 3). The table

TABLE 3
Freshman and Senior Scholarship of Students at Different IQ Levels

1st-semester college grades					7-semester college grades				
IQ	N	%	Mean	Range	Possible	Actual	%	Mean	Range
166-170	3	0.2	97.7	97-99	3	2	66.7	96.0	93-99
161-165	1	0.1	48.0	48	0				
156-160	3	0.2	67.3	32-92	0				
151-155	8	0.6	85.4	45-98	2	1	50.0	87.0	87
146-150	21	1.5	70.4	13-98	14	8	57.1	70.0	26-97
141-145	36	2.5	65.8	4-99	16	10	62.5	69.1	12-99
136-140	85	5.9	65.3	4-98	54	42	77.8	67.5	1-99
131-135	117	8.1	59.8	1-99	64	49	76.6	50.2	5-99
126-130	210	14.5	55.6	1-99	126	80	63.5	52.8	1-99
121-125	290	20.1	53.5	1-99	189	123	65.1	52.1	5-99
116-120	294	20.4	44.3	1-99	190	118	62.1	42.2	1-94
111-115	191	13.3	38.9	1-97	129	78	60.5	39.4	2-95
106-110	123	8.4	33.6	1-97	84	44	52.4	32.9	2-69
101-105	46	3.2	37.2	1-98	29	14	48.3	44.4	2-96
96-100	9	0.6	28.0	3-42	9	5	55.6	31.0	26-36
91-95	6	0.4	26.5	7-69	6	3	50.0	18.3	6-35
-90	1	0.1	27.0	27	0				
Total	1444	100.0			915	577	63.1		

also shows the range of scholarship at each level, and the proportion of those in college long enough to reach senior status at each level. (Scholarship is indicated by centiles.) The table shows that:

a. Although the general tendency is for those of higher IQ to make the better scholastic records, the range of scholastic performance is remarkably similar at the different levels. Achievement ranges between the lowest and the highest decile at each level from an IQ of 101 to that of 141. One student with an IQ of 105 attained a 96 centile standing as a senior. That the IQ was, however, an inaccurate index of her ability is suggested by the fact that she made a 71 centile score on the *OSU Test*.

b. Of 15 students with IQ's above 150, 12 made exceptionally good records.

c. Sufficient time had elapsed for 15 students whose IQ's are below 101 to become seniors. Eight of these have obtained the A B, but in only four cases was this achieved in the normal four-year period.

d. Some degree of selective persistence is evidenced by comparing the upper with the lower half of the distribution. Of the 468 with IQ's above 120, 315, or 67.3 per cent, persisted, as compared with 262, or 58.6 per cent, of the 447 below that point. The ratio of the difference in these proportions to the standard error of the difference is 2.74.

*Relationship between Aptitude and Achievement
at Different Levels of Motivation*

It is a matter of common observation that one of the basic reasons why the correlations between scholastic aptitude test scores and scholarship is no higher is that the degree of motivation varies widely from student to student. After comparing several formulas for computing an index of effort, Fei Tsao (1) has proposed the one which follows:

$$FQ = \frac{E}{\text{Predicted } \bar{E}} \times 100$$

in which "E" is an individual's scholastic grade score, and "Predicted E" is the value of the scholastic score which might be predicted from his aptitude test score when one employs the ordinary regression equation. If a student's grade average equals his predicted average, the value of the FQ is 100, this index representing normal effort; if his achievement is higher

than his predicted achievement, the value of the FQ is more than 100; and vice versa. This formula was used in computing Effort Quotients for the 1,444 population, using the IQ and *OSU Test* scores as independent and comparable bases of calculation, along with the high-school scholarship, for the purpose of predicting first-semester college scholarship, and for the 577 who had become seniors, calculating the FQ from the same two intelligence tests, but now employing first-semester scholarship for predicting scholarship for seven semesters.

Correlations between Aptitude Test Scores and First-Semester Scholarship at Different FQ Levels

Table 4 tabulates for each of the FQ levels the correlation

TABLE 4

Correlations at Different FQ Levels between Aptitude Test Scores and First-Semester Scholarship, when FQ is Computed from High-School Scholarship and (1) IQ, (2) OSU Test Scores.

FQ levels	r using IQ's	Scholarship centile		r using OSU	Scholarship centile	
		N	Mean		N	Mean
130 & above	.264	125	61.56	.338	101	51.33
120-129	.438	198	64.39	.610	155	56.68
110-119	.541	252	61.39	.692	304	64.67
100-109	.465	269	49.12	.642	292	50.34
90-99	.244	223	42.85	.483	199	42.34
80-89	.373	134	43.51	.549	164	46.77
70-79	.279	95	28.89	.469	97	29.22
60-69	.245	75	34.87	.518	51	37.55
Below 60	.145	73	18.97	.385	82	23.05
Total	.358	1444	50.00	.566	1444	50.00

between the aptitude test score and first-semester scholarship, together with the numbers and means, permitting a comparison between the results when two different aptitude indices are used, the IQ and the *OSU Test* scores. The predictive index in this case is computed from high-school scholarship.

These data indicate (1) that as the correlation between the IQ and first-semester grades is considerably lower than the corresponding correlation between scholarship and the *OSU Test* score (.358 as compared with .566) this therefore holds true at the different motivation levels. (2) There is consider-

able contrast between the size of r at the different levels. For the calculations employing the IQ, the range is from .145 to .541; for calculations employing the *OSU Test* score, the range is from .385 to .692. (3) The level where the relationship between test scores and scholarship is closest is 110-119. (4) The level where the students make the highest average grades is not that representing the highest FQ, but is one or two steps below the highest. (5) However, the lowest grades on the average are made by those with the lowest effort indices. (6) It is also at the level of low motivation where the correlation between aptitude and achievement is most out of step.

*Correlations between Aptitude Scores and Seven-Semester
Scholarship at Different FQ Levels*

Table 5 is comparable to Table 4, except that in the calcula-

TABLE 5

Correlations at Different FQ Levels between Aptitude Test Scores and Seven-Semester Scholarships, when FQ is Computed from 1st-Semester Scholarship and (1) IQ, (2) OSU Test Scores, with Mean Scholarship and Percentage of Persistence

FQ levels	r using IQ's	Scholarship centile		r using OSU	Scholarship centile		Percentage persisting
		N	Mean		N	Mean	
170-Over	.592	76	72.61	.477	53	60.03	74
150-169	.618	79	69.93	.630	55	61.86	76
130-149	.601	82	61.84	.763	74	63.61	70
110-129	.645	68	52.12	.716	112	67.38	78
90-109	.661	64	49.56	.633	78	45.37	67
70-89	.244	64	33.63	.512	75	34.83	60
50-69	.188	62	25.34	.458	58	27.90	59
30-49	.005	49	18.56	.262	46	19.85	53
Below 30	.015	33	11.56	.210	26	12.42	27
Total	.322	577	50.00	.494	577	50.00	63

tion of the FQ first-semester college scholarship, rather than high-school scholarship, is used in the prediction formula. Again a comparison is made of the correlations between aptitude test score and achievement (in this instance cumulative grades for seven semesters) at the different FQ levels. (1) the r 's again run considerably higher where the computation of FQ is based on the *OSU Test* than when the IQ is employed. (2) The FQ extends over a wider range at both extremes than is the case

when predicting first-semester grades, reaching from below 30 to over 170. (3) The range of r 's at the different motivation levels is even greater than was found in the case of the first-semester grades. Where the IQ is used it extends from .005 to .661, and where the *OSU Test* score is used in computing the FQ, the range is from .210 to .763. (4) The correlation between aptitude and achievement is again closest at a point below the highest level, although it is nearer the upper limit than is true in the case of freshmen. (5) Where the FQ is based upon the IQ there is a perfect positive correlation between degree of motivation and mean cumulative scholarship, but where the FQ is based upon the *OSU Test* score the highest grades were made by those with motivation indices of 110-129. From this high point the grade curve slopes regularly in both directions. (6) By the time the students with the lowest level of motivation reached senior status their grade average was a 12 centile. The literal grade equivalent of a 12 centile average for seven-semester cumulative scholarship is a low "C," whereas the letter equivalent of the 73 centile is "B." (7) Sufficient time had elapsed to make it possible for 915 out of the original freshman population of 1,444 to have become seniors. The last column in Table 5 indicates the proportion at each FQ level that persisted into the senior year—FQ being computed from *OSU Test* scores and first-semester scholarship. The largest proportion of those persisting occurred at the 110-129 level, where the per cent is 78. At the lowest motivation level the elimination is 73 per cent, and at the highest level it is 26 per cent. (The elimination represented by these figures is somewhat higher than normal; 38 men from the most recent classes had left college to enter the armed forces. Their FQs are, however, distributed over the population range in such chance fashion as not to disturb the general averages.) (8) For the total group the correlation between FQ and scholarship for seven semesters is .702, when FQ is computed from IQ, and .559, when FQ is based on *OSU Test* scores.

Relation between FQ and Phi Beta Kappa

For the purpose of discovering the relationship between FQ and the highest level of scholastic attainment, Table 6 was pre-

TABLE 6

Proportions at Decile Levels making Phi Beta Kappa when Students are Grouped according to 1) First-Semester Grades, 2) OSU Test Scores, 3) IQ's, and 4) FQ's (derived from OSU Test Scores and 1st-Semester Scholarship).

Decile levels	1st-se- mester grades N	PBK		OSU Test scores N	PBK		IQ's N	PBK		FQ's N	PBK	
		N	%		N	%		N	%		N	%
91-100	59	41	69	47	31	66	63	22	35	63	6	10
81- 90	79	18	23	52	11	21	65	11	17	65	7	11
71- 80	61	4	7	54	11	20	49	3	6	77	30	39
61- 70	65	1	2	71	5	7	62	16	26	61	15	25
51- 60	57	0	0	50	2	4	54	3	6	65	4	6
41- 50	72			49	1	2	75	4	5	59	2	3
31- 40	62			61			41	1	2	62		
21- 30	56			67	1	1	65	2	3	49		
11- 20	48			61	2	3	57	1	2	51		
1- 10	18			65			46	1	2	25		

pared to show the proportions elected to Phi Beta Kappa at different motivation levels, as compared with the proportions from equivalent decile levels of (1) first-semester scholarship, (2) *OSU Test* scores, and (3) IQ. Phi Beta Kappa election is usually granted to the upper eighth of the class.

Table 6 shows that final scholastic honors can be best predicted from freshman grades, 69 per cent having ranked in the highest tenth of the class as freshmen, whereas the corresponding figure for the *OSU Test* is 66 per cent, and for the IQ, 35 per cent. This might have been anticipated from the fact that the correlation between first-semester scholarship and cumulative seven-semester scholarship (which includes first-semester grades) is .805, whereas the validity coefficient for the *OSU Test* is .559, and for the IQ it is but .322. PBK election was obtained by a few students who ranked as low as the lowest deciles in the test distribution. In terms of FQ, the level contributing the largest proportion to PBK membership is 71-80, from which level 39 per cent were elected. This corresponds approximately to the 127-140 FQ group.

Summary

1. A total of 1,444 IQ scores are available for freshmen and 577 for seniors, these having been derived from the following tests: *Otis*, *Terman*, *Henmon-Nelson*, *National*, *Kuhlmann-*

Anderson, and *Stanford-Binet*. Each student was tested with the *OSU Psychological Examination*.

2. When the correlation figures are compared with those obtained with the smaller population, concerning whom a report was made in 1941, no significant differences appear.

3. The mean IQ of the Oberlin group is 122, the range: 90-169, and sigma: 10.7. The mean for those who became seniors was also 122.

4. Cumulative scholarship for seven semesters ranges from the lowest to the highest decile at each level from an IQ of 101 to 141, and 8 students achieved the A.B. whose IQ's were below 101, although in half of these cases this required more than eight semesters.

5. Some selective elimination on the basis of IQ is indicated by the fact that 67.3 per cent of those with IQ's above 120 became seniors, as compared with 58.6 per cent of those with IQ's below 121. The comparable figures for the upper and lower half of the *OSU Test* distribution are 68.9 and 57.3 per cent. The FQ is a still better basis for predicting persistence, with 73.6 of the upper half as compared with but 52.5 per cent of the lower half continuing into the senior year.

6. The correlation between aptitude test scores and freshman scholarship ranges from .145 to .541, when FQ is computed from IQ; from .385 to .692, when FQ is computed from *OSU Test* score. When the correlations are made with cumulative seven-semester scholarship, the range is still greater. Where the IQ is used it extends from .005 to .661; where *OSU Test* score is used it ranges from .210 to .763.

7. Although the lowest correlations and lowest scholarship correspond to the lowest FQ's, the highest correlations and highest achievement are found at levels somewhat below the highest FQ's.

8. It is the 127-140 FQ level which contributes the largest proportion of students elected to Phi Beta Kappa.

Conclusions

Data from this study confirm the results obtained in 1941 with a smaller population to the effect that it is possible for

a student with an IQ as low as 91-95 to obtain an A.B. in Oberlin College. The fact that this achievement is apt to require more than the usual number of semesters emphasizes the role of determination and directionality. Analysis of the relationship between motivation and achievement shows, however, that the highest grades are not usually made by those with the highest effort indices, but by those within the 56-85 centile range. Because of the method by which the FQ index is computed, the student with an IQ of 100, who makes average grades, obtains a higher index than the one with an IQ of 170, whose achievement is commensurate with his ability; but the former is much less apt to achieve Phi Beta Kappa distinction. The study also shows that when students with low effort indices are grouped, the aptitude test scores give little or no basis for predicting their scholarship, whereas with an able group, who are well motivated, the correlation may be quite high. In the case of the *OSU Test*, the difference is represented by r 's of .210 and .763.

REFERENCES

1. Tsao, Fei. "Is AQ or F Score the Last Word in Determining Individual Effort?" *Journal of Educational Psychology*, XXXIV (1943), 513-526.
2. Hartson, L. D. and Sprow, A. J. "The Value of Intelligence Quotients Obtained in Secondary School for Predicting College Scholarship." *Educational and Psychological Measurement* I (1941), 387-398.
3. Terman, L. M. and Merrill, M. A. *Measuring Intelligence*. Boston: Houghton Mifflin Company, 1937.

STATISTICAL INTERPRETATION OF SYMPTOMS ILLUSTRATED WITH A FACTOR ANALYSIS OF PROBLEM CHECK LIST ITEMS

STANLEY S. MARZOLF AND ARTHUR HOFF LARSEN

Illinois State Normal University

BEHAVIOR symptoms are often found to occur in conjunction more frequently than can be accounted for by chance. When the behavior symptoms are of the maladjustive sort it is customary to refer to their conjunction as a syndrome. Though belief in such syndromes is supported by clinical observation, this same sort of observation yields the conclusion that many cases of maladjustment cannot be considered typical of any recognized syndrome. Furthermore, observation reveals that it is difficult to specify an invariant conjunction of symptoms by which a syndrome may be defined. Elsewhere (3) it has been suggested that these situations and related problems can best be clarified by a statistical interpretation of symptoms. One statistical conception which is useful is factor analysis.

It is the purpose of this article to illustrate how syndromes may be conceived of in terms of factor analysis. Viewed in this way it will be seen that (1) atypical cases are not inconsistent with the recognition of syndromes, and (2) a syndrome is a "central tendency" concept which cannot be, and is not necessarily defined as, an invariant conjunction of symptoms.

The Mooney *Problem Check List* (4) was given to 205 upper classmen in Illinois State Normal University. This instrument consists of 330 items about which students might worry. In using the instrument the student is asked to underline those items about which he worries. When this has been done, he is further instructed to circle those items about which he is seriously concerned. The ten most frequently underlined items were intercorrelated. This was done with the aid of

computing diagrams (1) which provide an estimate of tetrachoric r . To have included more items would have involved cell entries so small as to make estimates grossly inaccurate. A direct centroid solution was made and this was transformed to orthogonal simple structure and to an oblique solution.

The extent to which the ten items were found to intercorrelate is shown in Table 1. Since the reliability of the check

TABLE 1
Intercorrelations among the 10 Most Frequently Underlined Items of the Mooney Problem Check List

	1 *	2	3.	4.	5.	6	7.	8.	9.	10.
1.*	..									
2.	.55	..								
3.	.38	.38	..							
4.	.37	.15	.40	..						
5	.28	.08	.40	.05	..					
6.	.16	.43	.24	.05	.18	..				
7.	.01	.02	.00	.15	.18	.23	..			
8.	.18	.05	.03	.22	.15	.08	.52	..		
9.	.04	.12	.08	.04	.12	.36	.25	.28	..	
10	.25	.00	.02	.04	.02	.07	.05	.12	.19	.

* The items are as follows

1. Wondering if I'll be successful in life
2. Wanting a more pleasing personality
3. Lack of self-confidence.
4. Afraid to speak up in class discussion
5. Disliking financial dependence on family
6. Taking things too seriously.
7. Not enough sleep
8. Too little chance to read what I like.
9. Not enough time for recreation.
10. Restless at delay in starting life work

list items is not known, but is probably low, all the intercorrelations are very likely attenuated. The standard error of an r , when the true correlation is zero, for an N of 205 is .069. Though the data are admittedly fallible, we do not believe that they are so much so as to prevent their use for the illustrative purposes we intend.

Though no negative correlations were obtained, it cannot be assumed that the true relationships are positive, owing to the unreliability of the obtained r 's.

The second factor residuals are shown in Table 2.

The B -group procedure, recommended by Holzinger and Harman (2, p. 23), did not produce clear-cut groups, but it

TABLE 2
Second Factor Residuals

	1 *	2	3.	4	5	6.	7.	8	9.	10.
1.*		.096	-.218	.015	-.007	-.112	-.056	.038	-.038	.170
2.	.096		-.034	-.105	-.133	.212	-.071	-.092	.023	-.071
3.	-.218	-.034		.077	.149	.022	-.003	-.046	.063	-.035
4.	.015	-.105	.077		-.118	-.123	.074	.104	-.040	-.017
5.	-.007	-.133	.149	-.118		.003	.055	-.004	-.006	-.046
6.	-.112	.212	.022	-.123	.003		-.027	-.196	.110	-.037
7.	-.056	-.071	-.003	.074	.055	-.027		.134	-.130	-.085
8	.038	-.092	-.046	.104	-.004	-.196	.134		-.089	-.017
9.	-.038	.023	.063	-.040	-.006	.110	-.130	-.089		.060
10.	.170	-.071	-.035	-.017	-.046	-.037	-.085	-.017	.060	

* Numbers refer to items in footnote to Table 1

was possible to arrange the items in a rough approximation of two groups. In Table 1 the first five items constitute one group and the last five a second. Since a group is defined as a set of items within a correlation matrix which correlate more with each other than with other items in the matrix, a group may be said to constitute a syndrome. How definitely the items of these groups should be thought of as a syndrome will be developed later after the results of the analysis are given. The fact that the items within a group show correlations with some items not in the group illustrates one reason why syndromes are not always clearly defined. If a greater number of items had been included in the matrix, the items which do not fit well in either of these two groups would probably be found to associate themselves with some of the added ones to constitute a third group or syndrome.

In order to illustrate the applicability of factor analysis to the interpretation of symptom intercorrelations, we turn now to these data. A direct centroid factor solution was obtained. Only two factors seemed to be of sufficient reliability, as judged on the basis of the size of the residuals, to merit consideration. The centroid factors were transformed to orthogonal simple structure, and the result is shown in Table 3.

By comparing the loadings of the items on the two factors, it will be seen that the factor axes pass through or very close to the points representing items three and seven (see Figure 1). In fact, the rotation from the centroid axes to orthogonal

TABLE 3
Orthogonal Simple Structure

Variable*	M ₁	M ₂	Communality
1.	.797	.101	.645
2.	.552	.142	.325
3.	.750	.001	.563
4.	.431	.118	.200
5	.335	.197	.151
6	.290	.405	.248
7.	.002	.633	.401
8	.101	.610	.382
9.	.022	.601	.361
10	.073	.213	.050
			3.326
Per cent of total variance	19.0	14.3	33.3

* Numbers refer to items in footnote to Table 1.

simple structure was determined by these two items. This procedure follows Reyburn and Taylor (5), who criticize Thurstone's criteria for obtaining simple structure. It is their contention that hypotheses concerning the probable nature of the factors should be used in determining the rotation, for to do so is just as defensible as to use hypotheses concerning the nature of factors as a basis for selecting items for intercorrelation. In the present instance, our choice was not one of complete freedom, for a meaningful analysis could not have been achieved by any arbitrarily chosen angle of rotation.

Consideration of the factor structure as shown in Table 3 indicates that items one to four are most closely related to the first factor and that items seven, eight, and nine are most closely related to the second factor. Thus one syndrome consists chiefly of confusion or worry about future success in life, a more pleasing personality, lack of self-confidence, and fear of speaking in class discussions. The second syndrome involves worry about lack of sleep, lack of opportunity to do preferred reading, and lack of opportunity to engage in recreation. These two syndromes will be recognized by counselors of college students as representing a sizeable proportion of those who come to their attention. The fact that the ten most frequently underlined items contain these two syndromes is thus in accord with expectation.

The three items, five, six, and ten, which are not so definitely characteristic of the two syndromes, would probably align themselves with other symptoms if the correlation matrix included more items. This is most likely true for items five and ten, since the total variance attributable to these two factors is small. It is of course possible that with a larger correlation matrix the groups here identified might not exist. However, all clinical observation makes this possibility seem unlikely. The magnitude of the loadings would doubtless change, but it seems likely that two such clusters would be found and that it would be possible to pass axes through them without doing violence to the data.

The distribution of the loadings on these two factors suggests the possibility that a bi-factor solution would be the best description of the data. However, since there were but two groups, such a solution was not made. It is doubtful if the solution, had it been made, would have added much to the interpretation, and at any rate the illustrative purpose of this discussion can be achieved without it.

We shall now turn to the question of causal factors. The first syndrome contains one item, namely, worry about lacking self-confidence, which is apparently more basic than the others in that it serves to identify the causal factor at the root of the syndrome. The decision to consider this as the causal factor responsible for the first syndrome comes from psychological knowledge other than that presented in these data. It is an inference based largely on case studies.

Inference from clinical observation is even more necessary if a possible causal factor responsible for the second syndrome is to be identified. Such observation suggests that the second syndrome may result from a lack of integration. Students are notoriously prone to run in all directions at once. A hierarchy of preferences is all too often absent with the result that students complain of not being able to study, read, or engage in recreation to the extent that they would like.

The possibility of correlated factors leads to a consideration of an oblique solution. The correlated factor axes were located so as to pass near item one and through the cluster

formed by items seven, eight, and nine. The angle of separation between these correlated reference vectors represents a correlation of .259. The location of the axes and their relation to the orthogonal reference vectors is shown in Figure 1. The

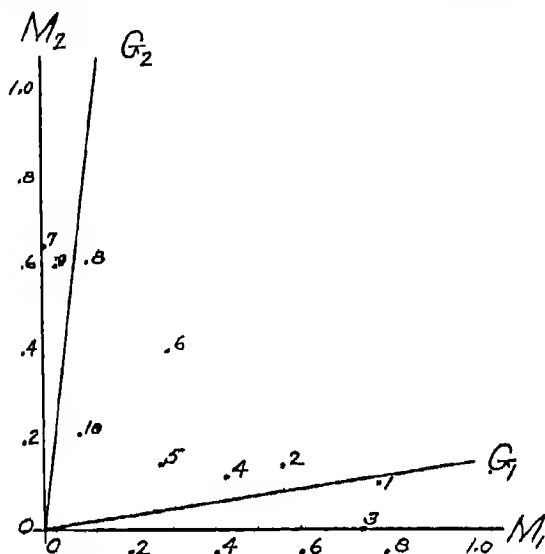


FIGURE 1
Orthogonal and Oblique Reference Vectors

correlated factor structure and pattern are presented in Table 4, while the contribution of the two correlated factors is given in Table 5.

TABLE 4
Correlated Factor Solution

Variable*	Structure		Pattern	
	r_1G_1	r_1G_2	G_1	G_2
1	.803	.198	.806	-.011
2	.566	.208	.549	.066
3	.744	.092	.772	-.107
4	.443	.169	.428	.058
5	.360	.236	.320	.153
6	.344	.438	.248	.374
7	.090	.629	-.078	.650
8	.185	.618	.027	.611
9	.106	.600	-.053	.614
10	.102	.220	.048	.208

* Numbers refer to items in footnote to Table 1.

The interpretation of the syndromes based upon the oblique solution does not differ greatly from that already discussed in relation to the orthogonal solution. There is, however, the problem of accounting for the correlation between these two correlated factors.

Several plausible bases for such correlation may be suggested. Students doubtless differ in their likelihood of worrying about anything. They may also differ in their willingness to admit worry. Ability to identify the source of their discontent may be another basis for differentiating students. Any one of these factors might be responsible for the correlation between the two vectors. Furthermore, lack of integration may be a factor in producing feelings of inadequacy, and thus the factors would be correlated. While the basis for correlation cannot be identified, there are a number of reasons why it might be expected to exist. In fact, a general factor of adjustment-nonadjustment, and hence correlated factors, are more to be expected than are uncorrelated factors.

TABLE 5
Total Contribution of Correlated Factors

	G_1	G_2
G_1	1.906	
G_2	.027	1.397
Grand Total 3.33		

Now let us suppose that all possible symptoms of maladjustment were intercorrelated and that all unreliability factors were absent. In this hypothetical situation, what might we expect factor analysis to reveal? It seems possible that a relatively small number of factors might be discovered, each representing an axis passing through a cluster of symptoms. These clusters would represent maladjustment syndromes. If symptoms of adjustment were also included in this hypothetical correlation matrix, we may conceive of bi-polar factors as representing the most adequate reference vectors.

Now one of the principal values to be attained by a consideration of symptoms in the statistical framework of factor

analysis is the manner in which it enables us to deal with the typical and atypical cases.

Consider first the orthogonal solution. Suppose that we have an individual with very marked feelings of inferiority or inadequacy, in other words one who "has" a large amount of the first factor. Since the second factor is uncorrelated with it, this individual may have much, none, or very little of the second factor. Assuming the factor loadings shown in Table 2 to be reliable, we would expect such a person to complain of worry about his future and the inadequacy of his personality, to be lacking in self-confidence, and to fear to speak up in class. He would be typical of the first or inferiority syndrome. Another individual might have no inferiority feelings but be greatly lacking in integration. This person would be typical of the second or unorganized syndrome. What is most likely, however, is that relatively more individuals would have moderate amounts of both factors, that is, would feel moderately inferior and be moderately lacking in integration. Such persons would not be typical of either syndrome and would show symptoms present in both. Thus it appears that we might expect that most people would not be "typical" cases. There remains the rare possibility that a given individual, on the basis of probability, would have large amounts of both factors. This person would not be typical either.

In the case of correlated factors the chief difference is that having a large amount of one factor would enhance the probability of having a greater-than-average amount of the other, a circumstance which would make a clear-cut distinction between syndromes even less likely than in the case of uncorrelated factors.

Factor analyses of intercorrelations between most frequently underlined items of the Mooney *Problem Check List* provide data for illustrative purposes. With these data the feasibility and desirability of factor analysis as a framework for clarifying clinical findings can be shown. It is demonstrated that though syndromes may be said to exist, it is not surprising that typical cases representing these syndromes are not more frequently encountered. Though these data deal

with the milder symptoms of maladjustment, the same principles may apply to the interpretation of psychoneuroses and psychoses.

In addition to making the existence of the atypical case consistent with the concept of syndrome, it is also possible to understand why syndromes can scarcely be defined as an invariant conjunction of symptoms. The magnitude of the correlations of the items of either of the syndromes with the factor assumed to be responsible for it is considerably less than perfect. Therefore, even though the basic factor is present we cannot expect that all of the symptoms will be. Various contingencies of the environment may exist without any relation to the existence of the causal factor and may thus alter its manifestation.

Furthermore, it is not inconsistent with the concept of syndrome for one symptom to be present in more than one syndrome; that is, to be moderately correlated with more than one causal factor. Item six, "taking things too seriously," may be an example of such a possibility.

A syndrome must be considered as a central tendency concept rather than an invariant conjunction of symptoms. Syndromes are recognized clinically because individuals with great amounts of the various causal factors do exist. These more dramatic instances are easily recognized. It is not inconsistent with their identification for us to find many other cases which are not typical.

The data also lend support to clinical observations regarding the frequency of certain behavior patterns in students. The evidence is most certainly not conclusive. The statistician will be inclined to doubt its validity and the clinician may feel that such evidence is unnecessary. On this matter we can only endorse Wolfe's (6, p. 40) statement that closer cooperation between clinician and factor analyst is desirable.

REFERENCES

1. Chesire, Leone, Saffir, Milton, and Thurstone, L. L. *Computing Diagrams for the Tetrachoric Correlation Coefficient*. Chicago: Chicago Medical Book Company, 1933.
2. Holzinger, Karl J. and Harman, Harry H. *Factor Analysis*. Chicago: University of Chicago Press, 1941.

3. Marzolf, Stanley S. "Symptom and Syndrome Statistically Interpreted." *Psychological Bulletin*, XLII (1945), 162-176.
4. Mooney, Ross L. *Problem Check List* Columbus: Bureau of Educational Research, Ohio State University, 1941.
5. Reyburn, H. A., and Taylor, J. G. "On the Interpretation of Common Factors: A Criticism and a Statement." *Psychometrika*, VIII (1943), 53-64.
6. Wolfe, Dael. *Factor Analysis to 1940*. Chicago: University of Chicago Press, 1940.

NEW TESTS OF 1944-1945

California Test of Personality Primary and Adult Series, by Louis P. Thorpe, Willis W. Clark and Ernest W. Tiegs, 1944. Scores are obtained for two main fields, self-adjustment and social adjustment. Each of these fields is further divided into six areas for which scores are obtained. Tests are in packages of 25, with a manual of directions. Primary series, Form A, per 25 tests, \$1.00, Primary series, Form B, per 25 tests, \$1.00, 5c per copy for tests when bought in smaller quantities, Primary specimen set, 25c. Adult series, Forms A and B, same prices as for Primary. Published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California.

ERC Stenographic Aptitude Test, by Walter L. Deemer, Jr., 1944. Time, 50 minutes. The purpose of this test is to estimate the probable performance of pupils studying shorthand. There are 5 subjects: speed of writing, word discrimination, phonetic spelling, vocabulary, sentence dictation. Package of 25, \$3.50, specimen set, 35c. Published by Science Research Associates, 228 S. Wabash Ave., Chicago 4, Illinois.

Foust-Schorling Test of Functional Thinking in Mathematics, by Judson W. Foust and Ralph Schorling, 1945. Time, 45 minutes. This test is intended to measure the power to deal with mathematical relationships. It is usable in the 9th grade and above. Sold only in packages of 25 copies with manual and key at \$1.35 per package, specimen set, 15c. Published by World Book Company, Yonkers-on-the-Hudson, N. Y.

General Clerical Test, 1944. Time, 50 minutes. This is a test which has been designed as a general and differential test for use in selecting persons for all types of clerical work. It is for high school and above. Sold in packages of 25 with manual and key. 1-3 packages, \$3.25 each, 4-39 packages, \$2.85 each; 40 or more packages, \$2.60 each. Per copy when package is broken, 15c. Specimen set, 30c. Published by the New York Psychological Corporation, 522 Fifth Ave., New York 18, N. Y.

Goldstein-Scheerer Test of Abstract and Concrete Thinking, by Kurt Goldstein and Martin Scheerer, 1945. These tests are for the use of psychologists, psychiatrists, and neurologists working with patients who have brain injuries. They are designed to measure both quantitatively and qualitatively the impairment of the function of the brain with reference to abstract and concrete reasoning. The monograph contains full instructions for the administration and interpretation of the tests. Three of the five tests are available: the *Goldstein-Scheerer Cube Test*, the *Weigl-Goldstein-Scheerer Color-Form Sorting Test*, and the *Goldstein-Scheerer Stick Test*. *Cube Test*: 2 design books, \$4.75 per set; Kohs blocks, \$1.00 per set of 12; record forms, sold only in packages of 50, \$2.75. *Stick Test*: 30 plastic sticks of 4 lengths, \$2.00 per set; record forms, including mimeographed supplement to the monograph, sold only in packages of 50, \$2.75. Complete materials, all three tests, including a copy of the monograph, "Abstract and Concrete Behaviour," \$21.50. Published by the Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

How Supervise? by Quentin W. Fife. No time limit. Norms are given in percentile rank and standardized scores. This test is considered as a measuring instrument of the supervisors' knowledge and insight concerning human relations in in-

dustry. It is designed to aid management in obtaining a clearer picture of its supervisors' understanding of the more important general aspects of the job. There are 3 sections: Supervisory Practices, Company Policies, and Supervisor Opinions. Forms A and B. Sold in packages of 25 with manual and key. 1 to 3 packages, \$2.00 each, 4 to 39, \$1.65 each, 40 or more, \$1.50 each, single copies, 10c. Specimen set, 30c. Published by The Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

Inventory of Factors GMIN (Abridged Edition), by J. P. Guilford and H. G. Martin. This test is designed to measure the following personality characteristics: G, general pressure for overt activity, A, ascendancy in social situations as opposed to submissiveness; M, masculinity of attitudes and interests as opposed to femininity; I, lack of inferiority feelings, and N, lack of nervous tenseness and irritability. Supplement to the manual and scoring keys are to be printed shortly, 100 copies, \$10.00, 50 copies, \$6.00, specimen set without scoring keys, 25c, specimen set with scoring keys, 50c. Published by the Sheridan Supply Company, P.O. Box 837, Beverly Hills, California.

Johnson Temperament Analysis, by Roswell H. Johnson, 1944. The form is self-administering and ordinarily requires 40 to 50 minutes. The following scores are obtained: nervous, depressive, active, cordial, sympathetic, subjective, aggressive, critical, and self-mastery. Specimen sets include test booklet, manual of directions, response record sheet, and 1 analysis profile sheet. Scoring stencils must be ordered separately. The price of the specimen set is 25c postpaid. Per 25 tests, \$1.25, 10c each in smaller quantities. Response record sheets, 1½c each, analysis profiles, 1c each. Published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California.

Mellenbruch Aptitude Test for Men and Women, by Paul L. Mellenbruch, 1944. Time, 35 minutes. This test is designed to discover the degree and limits of the mechanical trainability of men and women applicants for mechanical positions. It is also intended as a counseling aid for selecting junior and senior high school students who will profit from industrial arts training. Scoring key is included free with each order of 25 or more booklets. Published by Science Research Associates, 228 Wabash Ave., Chicago 4, Illinois.

Minnesota Multiphasic Personality Inventory, Group Form, by Starke R. Hathaway and J. Chamley McKinley. Time required is from 30 to 90 minutes. The individual form, published earlier, has been adapted to group form for use with those not requiring individual attention. For subjects from about 16 years of age upward. The following scores are obtained: hypochondriasis, depression, hysteria, psychopathic deviate, paranoia, masculinity, hypomania, validity, tendency to lie and tendency to give "cannot say" replies. Booklets may be reused. 1-24 booklets, 25c each, 25 or more, 22c each. Keys (either hand-scoring or machine-scoring should be specified) and manual, \$7.50. IBM answer sheets, \$5.00 per 100, 6c each in smaller quantities. Published by the Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

Occupational Interest Inventory—Intermediate, Form A, devised by Edwin A. Lee and Louis P. Thorpe, 1944. For junior high-school students to adults. The following scores are obtained: personal-social, natural, mechanical, business, the arts, the sciences, verbal, manipulative, computational, and interest level. Package of 25 tests with manual, \$1.75, 8c each in smaller quantities; specimen set, 25c. Published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California.

Occupational Preference Inventory, by Paul P. and Ralph T. Brainard, 1945. No time limit. This blank replaces the *Brainard and Stewart Specific Interest Inventories*, which will go out of print shortly. Occupational preferences are divided into 28 occupational sections, and the 28 sections are combined into 7 major occupational fields: commercial, personal service, agriculture, mechanical,

professional, esthetic and scientific. Booklets can be reused. Inventory booklets with manual 1 to 9 copies, 25c each, 10 to 99, 22c each, 100 or more, 20c each; Manual is 25c extra when fewer than 10 are ordered, record forms are sold only in packages of 50. 1 package is \$3.75, from 2 to 19, \$3.50 each, 20 or more, \$3.00 each; per copy, 10c each. Specimen set is 50c. Published by the Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

Personal Audit (Revised), Forms LL and SS, by Clifford R. Adams, 1945. No time limits. Form LL consists of 9 parts, each part containing 50 items. This form can be used with adults having the equivalent of a grammar school education or with senior high school students. The first six parts of Form LL are published separately as Form SS. Form SS is recommended for use with junior high school students or in other situations where testing time must be curtailed. Scoring requires no keys. The fields for which scores are obtained are seriousness-impulsiveness, tranquility-irritability, frankness-evasion, stability-instability, tolerance-intolerance, steadiness-emotionality, persistence-fluctuation, contentment-worry. Form LL package of 25 tests with manual, \$3.75, specimen set, 35c. Form SS package of 25 tests with manual, \$2.50, specimen set, 25c. Published by Science Research Associates, 228 S. Wabash Ave., Chicago 4, Illinois.

Pintner General Ability Tests. Non-Language Series, by Rudolph Pintner, 1945. These group tests are designed to measure mental functions independently of word knowledge and facility, and may even be administered in pantomime without the use of language, using special directions which can be obtained on request. The series is to consist of two batteries: the Intermediate Test for grades 4 through 9, and the Advanced Test for grades 9 and above. The Intermediate Test is now available. Package of 25 tests with manual, \$1.80, specimen set, 30c. Published by the World Book Company, Yonkers-on-the-Hudson, New York.

P-L-S Journalism Test, by George H. Phillips, Harry Levinson and H. E. Schrammel, 1944. Time, 40 minutes. This test is designed for use in high school and college classes which have completed a first course in journalism. Percentile norms are given. Per package of 25, directions and key included, \$1.00 f.o.b. Emporia, \$1.15 postpaid. In quantities less than 25, test 5c, directions 5c, specimen set, 15c. Published and distributed by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.

Reading Achievement Test—Intermediate Test. Forms A and B, by Donald D. Durrell and Helen Blair Sullivan, 1944. Time, 30 to 45 minutes. These forms of the *Reading Achievement Test* are designed for grades 3 to 6. Sold only in packages of 25 with manual and key. Form A or B, \$1.55. Specimen set, 55c. Published by the World Book Company, Yonkers-on-the-Hudson, New York.

Schrammel-Otterstrom Arithmetic Test, by H. E. Schrammel and Ruth Otterstrom, 1945. Time, 50 minutes. This test consists of two divisions. Test I for grades 4, 5, and 6, and Test II, for grades 7 and 8. Percentile norms are provided for interpreting the scores for each part and for the entire test, both mid-year and end-of-year testing. Per package of 25, directions and key included, 60c f.o.b. Emporia; 75c postpaid. In quantities less than 25, test 3c, key 3c, directions 5c. Specimen set, 20c. postpaid. Published and distributed by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, Kansas.

Stevens Reading Readiness Test, by Avis Coultas Stevens, 1944. These tests assist the teacher to group her beginning children for reading. They are not intended to displace intelligence or kindergarten tests. Package of 25 tests with manual and key, \$1.80; specimen set, 20c. Published by the World Book Company, Yonkers-on-the-Hudson, New York.

Survey of Mechanical Insight, designed by D. R. Miller, 1945. Time, 25 minutes. Designed to measure aptitude for solving the types of problems involved in jobs

requiring the operation, maintenance, repair, or design of various types of machinery. Package of 25 tests with manual and key, \$2.50, 11c each in smaller quantities. Published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California

Survey of Object Visualization, by D R Miller, 1945 Time, 20 minutes. This survey requires the examinee to predict how an object will look when its shape and position are changed. Tentative percentile norms are given. The test is practically self-administering and the scoring is completely objective. Package of 25 tests, including a manual of directions and scoring key, \$1.50; 7c a copy in smaller quantities. Published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California.

Survey of Space Relations Ability, by Harry W. Case and Floyd Ruch, 1944 Time, 15 minutes. This test is designed to measure the ability of the employee or applicant to perceive rapidly and accurately the relationships among objects in space. Tentative percentile norms are furnished. Package of 25 tests, including a manual of directions and scoring key, \$2.00; specimen set, 25c. Published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California

Wechsler Memory Scale, by David Wechsler, 1945. This a short individual test of memory arranged for use by psychologists working in mental hospitals. Sold only in packages of 50 record forms, with a set of cards for one of the tests, \$1.50, manual must be ordered separately, 45c, specimen set, including manual, design card and record form, 70c. Published by the Psychological Corporation, 522 Fifth Avenue, New York 18, N. Y.

Wellesley Spelling Scale, by Thelma G. Alper and Edith B. Mallory, 1944. This spelling ability test is for use at high school and college levels. Percentile norms are provided for the various grades from nine to college freshmen. Forms I and II, per 25 tests, 75c each; in smaller quantities, 4c per test. Published by the California Test Bureau, 5916 Hollywood Blvd., Los Angeles 28, California.

The United States Armed Forces Institute Tests of General Educational Development, 1945. The purpose of these tests is to measure the extent to which all of the past educational experiences of the individual tested have contributed to his general educational development or to his ability to carry on an educational program in high school or in the first two years of college. Emphasis is placed on intellectual powers rather than detailed content. Cost of the tests may be had upon application to the distributors. Published by the American Council on Education. Distributed by the Cooperative Test Service of the American Council on Education, 15 Amsterdam Ave., New York 23, N. Y., and Science Research Associates, 228 S. Wabash Ave., Chicago 4, Illinois.

High School Level Tests. These tests are to determine whether the individual has had the equivalent of a high school education. Standard scores are provided, percentile norms for the total score are available. The tests include:

- Test 1. *Correctness and Effectiveness of Expression.*
- Test 2. *Interpretation of Reading Materials in the Social Studies.*
- Test 3. *Interpretation of Reading Materials in the Natural Sciences.*
- Test 4. *Interpretation of Literary Materials.*
- Test 5. *General Mathematical Ability.*

College Level Tests. These tests are to determine whether the individual is capable of carrying on advanced college work. Standard scores are provided; percentile norms are available. The tests include.

- Test 1. *Correctness and Effectiveness of Expression.*
- Test 2. *Interpretation of Reading Materials in the Social Studies.*
- Test 3. *Interpretation of Reading Materials in the Natural Sciences.*
- Test 4. *Interpretation of Literary Materials.*

Subject Tests These tests are to determine the individual's proficiency in specific subjects on both high school and college levels. The tests include

Examination in English—High School Level; Book I: Reading and Interpretation of Literature and Literary Acquaintance. Percentile norms for total score available for grades 10, 11 and 12.

Examination in English—High School Level, Book II: Composition Norms are available

Examination in English—College Level, Book I: Reading and Literary Acquaintance Percentile norms for total score are available

Examination in English—College Level, Book II: Composition Percentile norms are available

Examination in Business English—High School Level. No norms available

Examination in Commercial Correspondence—College Level. No norms available

Examination in French Vocabulary—Lower Level No norms available

Examination in French Reading Comprehension—Lower Level. No norms available

Examination in French Grammar—Lower Level. No norms available

Examination in French Vocabulary—Upper Level No norms available.

Examination in French Reading Comprehension—Upper Level No norms available

Examination in French Grammar—Upper Level No norms available

Examination in German Vocabulary—Lower Level No norms available

Examination in German Reading Comprehension—Lower Level. No norms available

Examination in Spanish Vocabulary—Lower Level. No norms available.

*Examination in Spanish Reading Comprehension—Lower Level—*No norms available

Examination in Spanish Grammar No norms available.

Examination in Italian Vocabulary No norms available

Examination in Italian Reading Comprehension—Lower Level No norms available.

Examination in Italian Grammar—Lower Level No norms available

Examination in Business Arithmetic—High School Level No norms available.

Examination in Advanced Arithmetic—High School Level No norms available

Examination in Elementary Algebra—High School Level Percentile norms for total score available

Examination in Second-year Algebra—High School Level. Percentile norms for total score available

Examination in Plane Geometry—High School Level. Percentile norms for total score available

Examination in Algebra—College Level. No norms available.

Examination in Plane Trigonometry—College Level No norms available

Examination in Analytic Geometry—College Level No norms available.

Examination in General Science—High School Level No norms available

Examination in Biology—High School Level. No norms available

Examination in Physics—High School Level. No norms available.

Examination in Chemistry—High School Level. Percentile norms for total score available

Examination in Meteorology—High School Level. No norms available

Examination in Senior Science—High School Level. No norms available

Examination in Chemistry—College Level. Percentile norms for total score available.

Chemistry Examination in Qualitative Analysis. No scaled scores provided. Percentile norms available

Chemistry Examination in Quantitative Analysis Percentile norms available

Examination in Organic Chemistry Percentile norms available.

Examination in Astronomy—College Level. No norms available

Examination in Elementary Psychology—College Level. No norms available

Examination in American History—High School and College Levels. Percentile norms available

- Examination in World History—High School Level* No norms available.
Examination in Civics—High School Level. No norms available
Examination in Problems of Democracy—High School Level. No norms available
Examination in German Grammar—Lower Level No norms available.
Examination in Business English—High School Level No norms available.
Examination in Commercial Correspondence—College Level No norms available
Examination in Business Arithmetic—High School Level. Percentile norms available
Examination in Gregg Shorthand—First Year—Secondary School No norms available
Examination in Typewriting—First Year—Secondary School. No norms available
Examination in Typewriting—Second Year—Secondary School No norms available
Examination in Bookkeeping and Accounting—First Year—Secondary School No norms available.
Examination in Bookkeeping and Accounting—Second Year—Secondary School No norms available.
Examination in Mechanical Drawing—High School Level No norms available.
Examination in Engineering Drawing—College Level No norms available
Examination in Machine Design—College Level No norms available.
Examination in Strength of Materials—College Level No norms available
Examination in Surveying—College Level. No norms available
-

THE CONTRIBUTORS

Herbert A. Toops is professor of psychology at the Ohio State University. He is the author of numerous articles and books in the fields of personnel, guidance, measurement, statistics and motivation. The *Ohio State University Psychological Examination*, developed by Dr. Toops, is used in the statewide high school testing programs of four states. During the war Dr. Toops spent considerable time in Washington. For six months of 1944 he was with the Office of the Secretary of War working on the development of a statistical reporting system for civilian employees of the War Department. He also spent three months in the early part of 1945 on a survey of the National Roster of Scientific and Specialized Personnel. He is a member of the editorial board of *Educational and Psychological Measurement*.

Milton M. Mandell is chief of the administrative and management testing sub-unit of the United States Civil Service Commission. His previous positions have included those of personnel officer, Materials Division, W.P.B.; assistant chief of examinations, Los Angeles City Civil Service Commission, and classification consultant, Connecticut State Personnel Department. Mr. Mandell was co-author with Wallace Sayre of "Education and the Civil Service in New York City," published by the United States Office of Education.

Major Thomas W. Harrell received his Ph D degree from the Johns Hopkins University in 1936. He joined the staff of the University of Illinois in 1936 and taught there until May, 1940. From 1938-1940 he conducted test research and consulted with the Air Corps Technical School, U. S. Army. From May, 1940, to June, 1942, he was Executive, later Chief, Personnel Research Section, A G O, War Department. Called to active duty in the Army in January, 1942, he continued in the Personnel Research Section until September, first in his former position and later as Air Liaison Officer from the A.G.O. to the Army Air Forces. In September he transferred to Headquarters of the A.A.F. where he was Classification Officer until May, 1944. Since that time he has been overseas serving as Air Liaison Officer, Hq. Replacement Command, NATOUSA until February, 1945, as Chief, Personnel Classification Audit, AAF, MTO, July-October, 1944, and then with Hq. 15th Air Force as Director, Manning Section, A-1 Division from February-July, 1945. He is now Deputy, Asst. Chief of Staff, Personnel, A-1, Hq. 15th Air Force.

For five years after graduation from Stanford University in 1933, Margaret S. (Mrs. Thomas W.) Harrell was research assistant for

her father, Dr. E. K. Strong, Jr., in his Vocational Interest Research Office. From 1939-1940 she was employed in the Research and Analysis Unit of the San Francisco Employers Council. In November, 1940, she transferred to the War Department where she was employed until June, 1942, first as a junior personnel technician in the Personnel Research Section, A.G.O., second in the Planning Division, Special Service, and third as chief clerk of the Morale Research Division, Special Service.

J. R. Wittenborn is assistant professor and clinical psychologist at Yale University. He divides his time between the department of psychology, where he gives graduate instruction in statistics and clinical psychology, and the department of university health, where he is responsible for the educational occupational orientation of undergraduates. Since May, 1944, he has been participating in a Committee on Medical Research project for the Office of Scientific Research and Development. He received his Ph.D. from the University of Illinois in 1942.

Charles I. Mosier is chief of research and test construction of the State Technical Advisory Service, Social Security Board. He was on the faculty of the University of Florida from 1933 through 1941 as instructor, assistant professor, assistant university examiner and acting university examiner. Prior to his present position with the State Technical Advisory Service, he was personnel research technician, chief of classification, and chief of personnel standards in the same organization. He received his Ph.D. in psychology at the University of Chicago in 1937. He is on the editorial boards of *Psychometrika* and *Educational and Psychological Measurement*, and is a member of the Southern Regional Committee, Social Science Research Council.

M. Claire Myers, after receiving her Ph.D. in psychology from the University of California, left the Institute of Child Welfare in Berkeley, where she had been in charge of research for the Child Guidance Study, to spend two years as a psychology instructor at Wellesley College. Since that time she has been with the State Technical Advisory Service of the Social Security Board. After spending one year as personnel consultant in the regional office for the Pacific Coast States, she came to her present position of personnel research technician in the Research and Test Construction Section.

Helen G. Price received an M.A. in psychology from Stanford University and is at present employed as personnel research technician in the State Technical Advisory Service of the Social Security Board. She has been a training supervisor in the Ohio Bureau of Unemployment Compensation, and has assisted in the research work of the Employment Stabilization Research Institute and in the standardization of Forms M and L of the Stanford Binet.

L. D. Hartson is professor and head of the department of psychology at Oberlin College. His duties have included directing the

Bureau of Tests and Measurements of the College. Dr. Hartson has published many articles reporting his studies on the validation of psychological tests and rating scales as well as the construction and evaluation of vocational interest questionnaires. Much of his time in recent years has been devoted to a study of the vocational histories of Oberlin graduates.

Stanley S. Marzolf is associate professor of psychology at Illinois State Normal University. Dr. Marzolf teaches graduate courses in the field of clinical psychology, and directs the graduate work of students planning to become guidance and personnel workers in high schools. His several activities in the testing and guidance fields include chairmanship of the research committee which administers the university testing programs. He received his Ph D in psychology from the Ohio State University in 1937.

Arthur Hoff Larsen is assistant dean and head of the department of education and psychology at Illinois State Normal University. He is also director of the interviewing and testing done in the Veterans Administration Guidance Center located on the campus. He received his Ph D. from the University of Wisconsin in 1939.

•

MEMBERS OF THE AMERICAN COLLEGE PERSONNEL ASSOCIATION

- Adams, Dr Clifford R., Marriage Counseling Service, Pennsylvania State College, State College, Pennsylvania
- Adams, Lt. Wm M., AGO Reconditioning Unit, Annex IV, Brooke General and Convalescent Hospital, Fort Sam Houston, Texas
- Aiken, Prof N J., Director, Placement Bureau, and Associate Professor of Economics, State College of Washington, Pullman, Washington
- Alexander, Miss Ruth Lenore, 804½ E. 8th Street, Bloomington, Indiana
- Albright, Miss Thelma, Dean of Students, Queens College, Charlotte, North Carolina
- Allen, Mr Cecil H., Head of the Division of Personnel Service, N E Missouri State Teachers College, Kirksville, Missouri
- Anderson, Mr Gordon V., Acting Director, Student Counseling Bureau, 101 Eddy Hall, University of Minnesota, Minneapolis, Minnesota
- Andrews, Dr Elizabeth G., Director of Personnel, Florida State College for Women, Tallahassee, Florida
- Applegate, Mr C William, 7740 N Ashland, Chicago 26, Illinois
- Austin, Lt Sidney F., Quarters C H, U S Naval Training Station, Newport, Rhode Island
- Babcock, Miss Fern, National Student Secretary, Y W. C. A., 600 Lexington Avenue, New York, New York
- Bagley, Miss Leila, 1000 Morningside Drive, New York 27, New York
- Bailey, Prof Harold W., Director, Personnel Bureau, University of Illinois, Urbana, Illinois
- Bathurst, Dr J E., Department of Psychology and Education, Birmingham-Southern College, Birmingham, Alabama
- Belknap, Miss A Fredericka, Director, Personnel Bureau, New Jersey College for Women, New Brunswick, New Jersey
- Bell, Capt Hugh M., Classification and Counseling Section, AGO 1E936 Pentagon, Washington, D C.
- Berg, Mr Irwin A., Clinical Counselor, and Associate in Psychology, University of Illinois, Urbana, Illinois
- Bergstresser, Dr John L., Dean of Men, College of the City of New York, Convent Avenue and 139th Street, New York, New York
- Biglow, Miss Mary D., Chairman, Advising, Stephens College, Columbia, Missouri
- Billings, Mr Earle Monroe, Personnel Director, Eastman Kodak Company, 343 State Street, Rochester, New York
- Blaesser, Dean Willard W., Assistant Dean of Students, University of Chicago, Chicago, Illinois
- Blanchard, D (Miss) Leslie, Principal, Holmquist School, New Hope, Pennsylvania
- Bishop, Mrs Joan Fiss, Director, Placement Office, Wellesley College, Wellesley 81, Massachusetts
- Blanding, Dean Sarah G., New York State College of Home Economics, Cornell University, Ithaca, New York
- Boyles, Lt R. E., 532 River Street, Hoboken, New Jersey
- Bradshaw, Dean Francis F., University of North Carolina, Chapel Hill, North Carolina
- Bridgman, Mr Donald S., American Telephone and Telegraph Co., 195 Broadway, New York, New York
- Bridgman, Pres. Ralph B., Hampton Institute, Hampton, Virginia

- Brooks, Mrs Dorothy, Dean of Women, Dennison University, Granville, Ohio
- Brottemarkle, Prof Robert A., Personnel Officer, 116 College Hall, University of Pennsylvania, Philadelphia, Pennsylvania
- Brouwer, Mr Paul J., 6658 University Avenue, Chicago 37, Illinois
- Brown, Mr Charles K., Director of Student Affairs and Publications Editor, Alabama Polytechnic Institute, Auburn, Alabama
- Brown, Mrs Lucile B., American Red Cross, APO 763, New York, New York
- Brumbaugh, Dr A J., American Council on Education, 744 Jackson Place, Washington 6 D. C.
- Brummett, Mr. Carl A., Director of Student Activities, Jefferson College, 1528 Locust Street, St. Louis, Missouri
- Bunn, Mr John W., Dean of Men, Stanford University, California
- Butler, Lt. John, A.S.T.P., Camp Wheeler, Georgia
- Cameron, Mrs Marian Lippitt, University Testing Service, Syracuse University, Syracuse, New York
- Camp, Miss Frances M., Director, Bureau of Appointments, State University of Iowa, Iowa City, Iowa
- Cannon, Dean C W., Consulting Psychologist, Stevenson, Jordon and Harrison, 205 W Wacker Drive, Chicago 6, Illinois
- Carpenter, Miss M Helen, Acting Director, Placement Bureau, University of Colorado, Boulder, Colorado
- Chambers, Mr O R., Head, Department of Psychology, Oregon State College, Corvallis, Oregon
- Clark, Miss Barbara M., Assistant Director, Student Activities Bureau, University of Minnesota, Minneapolis, Minnesota
- Coates, Miss Dorothy R., Director, Katharine Gibbs School, 720 N. Michigan Avenue, Chicago, Illinois
- Colvin, Mr Harold William, Y M C A National Council, 19 S. LaSalle Street, Chicago, Illinois
- Compton, Dean Richard K., Hastings, Michigan
- Congdon, Dean Wray H., Dean of Undergraduates, Lehigh University, Bethlehem, Pa.
- Congdon, Dr Nora A., Personnel Research Officer, Colorado State College of Education, Greeley, Colorado
- Conwell, Dean H H., Beloit College, Beloit, Wisconsin
- Corson, Mr James Hunt, Dean of Men, College of the Pacific, Stockton, California
- Crenshaw, Mrs. James L., Director, Bureau of Recommendations, Bryn Mawr College, Bryn Mawr, Pennsylvania
- Croft, Lt. Col Lysle W., Hdq. Inf., W D Personnel Center, Camp Atterbury, Indiana
- Crawley, Dr. S. L., Director, Department of Student Personnel, Colorado State College of Education, Greeley, Colorado
- Crosby, Richard C., Director of Student Counseling, Miami University, Oxford, Ohio
- Cunnigim, Miss Margaret L., Dean of Women, Ripon College, Ripon, Wisconsin
- Darley, Lt. (jg), H-V (S) USNR, Bureau of Medicine and Surgery, Aviation Medical Division, Navy Dept., Potomac Annex, Washington 25, D C
- Dalton, Glen D., Y M C A., 600 Lexington Avenue, New York, New York
- Davies, Mr. Roy O., Director of Personnel, Western Auto Supply Co., 2107 Grand Avenue, Kansas City, Missouri
- Davis, Mr. Victor M., Director, Bureau of Appointments, University of Tennessee, Knoxville, Tenn.
- Dawson, Mr. J. D., Personnel Director, Antioch College, Yellow Springs, Ohio
- Deabler, Dr Herdis L., Personnel Director, North Central College, Naperville, Illinois
- DeGolyer, Mrs. Robert S., 575 Madison Avenue, S.E., Grand Rapids, Michigan
- Densmore, Mr. Harvey B., Chairman of General Studies, University of Washington, Seattle, Washington
- DeVoss, Dr. James C., Dean, Upper Division, San Jose State College, San Jose, California

- Dillman, President Grover C, Michigan College of Mining and Technology, Houghton, Michigan
- Donahue, Dr Wilma T, Department of Psychology, University of Michigan, Ann Arbor, Michigan
- Doty, Miss Katherine S, Assistant to the Dean, Barnard College, Columbia University, New York, New York
- Dozier, Prof. Doris, Recorder and Secretary of Placement, Mills College, Mills College, California
- Dressler, Miss Marguerite R, Assistant Director of Personnel, Florida State College for Women, Tallahassee, Florida
- Drought, Dr. Neal E, RCA Victor Division, Personnel Administration, Camden, New Jersey
- Dugan, Mr Willis E, Director of Student Personnel, University High School, University of Minnesota, Minneapolis, Minnesota
- Dunham, Mr. Charles Vernon, Director of Student Employment, University of Texas, Austin 12, Texas
- Dysinger, Dr. Wendall S, Dean, MacMurray College, Jacksonville, Illinois
- Edwards, Miss Marcia, Professor of Education and Assistant to Dean, College of Education, University of Minnesota, Minneapolis, Minnesota
- Eldersveld, Miss Wilma, Psychological Clinic, 1027 E Huron Street, University of Michigan, Ann Arbor, Michigan
- Elias, Miss Virginia, Director, Student Aid and Employment, Occidental College, 1600 Campus Road, Los Angeles 14, California
- Ellingson, President Mark, Rochester Institute of Technology, 65 South Plymouth Avenue, Rochester 8, New York
- Esterly, Mrs Virginia Judy, Assistant to the President, Scripps College, Claremont, California
- Evans, Mr John J, Jr, Armstrong Cork Company, Lancaster, Pennsylvania
- Evans, Miss Mary Catherine, Vocational Adviser for Women, Indiana University, Bloomington, Indiana
- Ewan, Mrs Mary A., Associate Chairman, Appointments Division, College of Education, Arps Hall, Ohio State University, Columbus, Ohio
- Failing, Miss Jean, New York State College of Home Economics, Cornell University, Ithaca, New York
- Falvey, Miss Frances, Adviser to Freshmen, Hollins College, Virginia
- Faries, Miss Miriam, Assistant Dean of Students, College of the City of New York, Convent Avenue and 139th Street, New York, New York
- Farrington, Mr. E H., Director, Placement Service, North Texas State Teachers College, Denton, Texas
- Feder, Lt Cdr Daniel D, Bureau of Naval Personnel, Training Division, 4721 Arlington Annex, Arlington, Virginia
- Feldt, Miss Irene C, Director, Placement for Women, Purdue University, Lafayette, Indiana
- Felsted, Mrs Leona Wise, Chairman, Personnel Committee, Illinois Wesleyan University, Bloomington, Illinois
- Fisher, Prof Charles A, Director, Education Placement Bureau, Temple University, Broad and Montgomery, Philadelphia, Pennsylvania
- Fisk, Miss Helen G, Associate Director, Western Personnel Service, 30 North Raymond Avenue, Pasadena, California
- Flood, Mr. Julian J., Personnel Director, Tuskegee Institute, P. O Box 445, Tuskegee, Alabama
- Focht, Miss Helen E., Old Capitol, State University of Iowa, Iowa City, Iowa
- Foley, Mr. John D., Assistant to the Dean, Office of Dean of Students, University of Minnesota, Minneapolis 14, Minnesota
- Gaetje, Miss Lucille, Director of Personnel for Women, Owens-Illinois Glass Company, Gas City, Indiana
- Gallagher, Prof Marie K, Chairman, Bureau of Educational Guidance, Hunter College, 695 Park Avenue, New York, New York
- Garrett, Mrs. Mary, Director of Admissions and Student Personnel, Bennington College, Bennington, Vermont
- Gebauer, Miss Dorothy, Dean of Women, University of Texas, Austin 12, Texas

- Gerken, Mr Clayton d'A., Department of Psychology, State University of Iowa, Iowa City, Iowa
- Gerth, Mrs. Josephine Hammond, Hunter College, 695 Park Avenue, New York, New York
- Gilbert, Mr William M., Clinical Counselor, Personnel Bureau, University of Illinois, Urbana, Illinois
- Goetch, Mr E W., Director, Placement Bureau, Iowa State Teachers College, Cedar Falls, Iowa
- Gordy, Miss Margaret D., Placement Director, Wilson College, Chambersburg, Pennsylvania
- Gregory, Lt Wilbur S, 0-582761, Box 731, 326th A.A F B U, MacDill Field, Florida
- Griffen, Miss Alice J., Director of Personnel Service, Wright Junior College, 3400 N Austin Avenue, Chicago, Illinois
- Griswold, Miss Mary Lou, Placement Secretary, New York State College of Home Economics, Cornell University, Ithaca, New York
- Gum, Miss Wanda N., Assistant Dean of Women, Southern Illinois Normal University, Carbondale, Illinois
- Gus, Mr Charles Edward, Executive Secretary, College of Engineering, New York University, University Heights, New York, New York
- Haggerty, Mr. William J., President, State Teachers College, New Paltz, New York
- Hale, Dr Lincoln B., President, Evansville College, Evansville 4, Indiana
- Hand, Mr Thomas J., Consultation Service, 87 Madison Avenue, New York 16, New York
- Hanson, Miss Anna M., Simmons College, Boston, Massachusetts
- Hanson, Mr Ernest E., Dean of Men, Northern Illinois State Teachers College, De Kalb, Illinois
- Hausam, Miss Winifred M., Director, Western Personnel Service, 30 North Raymond Avenue, Pasadena, California
- Held, Lt Cdr. Omar C., (S), USNR, Navy V-12 Unit, Colgate University, Hamilton, New York
- Helser, Dean Maurice David, Personnel Director, Iowa State College, Ames, Iowa
- Henderson, Mrs. Virginia Kinsman, Personnel Officer, 19 West Plumstead Avenue, Lansdowne, Pennsylvania
- Henry, Prof. Edwin Ruth van, A.G O, 270 Madison Avenue, New York, New York
- Henley, Dr. W. Ballentine, President, College of Osteopathic Physicians, 1721 Griffen Avenue, Los Angeles, California
- Hicks, Mr. Arthur C., Acting Registrar, Western Washington College of Education, Bellingham, Washington
- Hill, Prof J Lawrence, Jr., Department of Mechanical Engineering, University of Rochester, Rochester, New York
- Hilliard, Mr George, Director of Personnel and Guidance, Western Michigan College of Education, Kalamazoo, Michigan
- Hill, Mr. George E., Director, Student Personnel Service, Macalaster College, St. Paul, Minnesota
- Hilton, Miss M Eunice, Dean of Women, Syracuse University, Syracuse, New York
- Hoekje, Mr. John E, Registrar, Western Michigan College, Kalamazoo, Michigan
- Holcomb, Miss Leila L, Assistant Counselor, University Residences for Women, University of Texas, Austin 12, Texas
- Holmes, Mr. John L, Assistant Director of Personnel, Iowa State College, Ames, Iowa
- Holmes, Dr. Lulu H., Dean of Women, State College of Washington, Pullman, Washington
- Hopkins, Lt. (jg) Everett H., D-V (S), U. S. N R., Dept of Personnel, U. S. Naval Training Station, Farragut, Idaho
- Hoppock, Prof Robert, Department of Education, New York University, New York, New York
- Humber, Dr Wilbur J, Dean of Student Affairs, Kalamazoo College, Kalamazoo, Michigan
- Humphreys, Dr. J Anthony, Director, Personnel Service, and Registrar, Woodrow Wilson Junior College, 6850 Stewart Avenue, Chicago, Illinois

- Ingalls, Les W., Acting Director, Admissions and Personnel for Men, Middlebury College, Middlebury, Vermont
- Johnson, Captain Albert P., C I S, Midland Army Air Field, Midland, Texas
- Johnson, Mr Clyde Sanfred, 202 Administration Building, University of California at Los Angeles, 405 Hilgard Avenue, Los Angeles 24, California
- Johnson, Prof William C., Palmyra College, Diamond, Ohio
- Jones, Dr Edward S., Director Personnel Research, University of Buffalo, Buffalo, New York
- Jones, Miss Katherine J., Executive Secretary, Nursery Training School of Boston, 355 Marlborough Street, Boston 15, Massachusetts
- Jones, Dr Lonzo, Dean of the Faculty, Central Missouri State Teachers College, Warrensburg, Missouri
- Jox, Mr Marshall J., Personnel Director, Valparaiso University, Valparaiso, Indiana
- Kennon, Miss Ann Byrd, Personnel Director for Women, University of Colorado, Boulder, Colorado
- Kirkpatrick, Mr Forrest H., Personnel Administration, Radio Corporation of America, Camden, New Jersey
- Kittrell, Miss Flemma P., Dean, Home Economics Division, Howard University, Washington 1, D C
- Kopas, Mr Joseph S., Director, Department of Development, Fenn College, Cleveland, Ohio
- Lane, Mrs Estella Hitchcock, Dean of Students, Frances Shimer College, Mount Carroll, Illinois
- Larsen, Dr Robert P., Outlook Sanatorium, Urbana, Illinois
- Lauer, Mr George N., Dean of Men, Central Michigan College of Education, Mt Pleasant, Michigan
- LaBarre, Miss Corrine G., Research Assistant, Western Personnel Service, 30 N. Raymond Avenue, Pasadena, California
- Leo, Brother I., St Mary's College, Winona, Minnesota
- Lahr, Miss Patricia A., Director, Student Union, University of Nebraska, 14th and R, Lincoln, Nebraska
- Leetch, Miss Florence K., Appointments Office, Harvard University, Cambridge 38, Massachusetts
- Leetch, Mr. George N P., Director, College Placement Service, The Pennsylvania State College, State College, Pennsylvania
- Lemon, Mr Erwin B., Dean of Admissions, Oregon State College, Corvallis, Oregon
- Lloyd-Jones, Prof Esther, Associate Director of Personnel, Teachers College, Columbia University, New York, New York
- Lyman, Mr E., Chairman, Board of Personnel Administration, Northwestern University, 102 Lunt Building, Evanston, Illinois
- MacKenzie, Mrs. Barbara O'Neil, Personnel Associate, Brooklyn College, Brooklyn, New York
- Mackenzie, Mr Donald M., North Central Association of Colleges and Secondary Schools, 5835 Kimbark Avenue, Chicago, Illinois
- MacDonald, James W., Director of Personnel, Wilmington College, Wilmington, Ohio
- Madsen, Mr. Howard C., Manager Technical Employment and Training, Westinghouse Electric and Manufacturing Company, 306 Fourth Avenue, East Pittsburgh, Pennsylvania
- Mallay, Miss Betty Anne, Director of Appointment Bureau, Manhattanville College of the Sacred Heart, New York, New York
- Manson, Dr Gracc E., Adjutant General's Office, Beekman Tower, 1st Avenue and 44th Street, New York, New York
- Manzer, Mr Charles W., Faculty Adviser, Washington Square College, New York University, New York, New York
- Marks, Mr Arlyn, Personnel Secretary, Comptroller's Office, University of Illinois, Urbana, Illinois
- Marshall, Dr Thomas O., Director of Student Personnel, Colorado State College, Fort Collins, Colorado
- McCall, Prof William C., Department of Education, University of South Carolina, Columbia, South Carolina

- McCarn, Mrs Ruth O, Counselor to Women, Northwestern University, Evanston, Illinois
- McCaul, Miss Margaret E, 448 E. Mississippi, Liberty, Missouri
- Mason, Mr. Harry M, Testing and Guidance Program, University of Texas, Austin 12, Texas
- McClintock, Lt. James A, U.S.N.R., Villanova College, Villanova, Pennsylvania
- McCreary, Dr. Otis C., 5151 Alcoa Avenue, Los Angeles, California
- McCurley, Miss Mary T., Assistant to the President, Goucher College, Baltimore, Maryland
- McKinney, Prof. Fred, University of Missouri, Columbia, Missouri
- Miller, President Ernest, Goshen College, Goshen, Indiana
- Mills, Miss Thelma, Director of Student Affairs for Women, University of Missouri, Columbia, Missouri
- Miner, Mr. Robert J, Assistant Dean of Students, College of the City of New York, Convent Avenue and 139th Street, New York, New York
- Morgan, Miss Margaret, Head, Division of Teaching Appointments, Appointments Office, Columbia University, New York 27, New York
- Moore, Mr. Robert F., Secretary of Appointments, Alumni House, Columbia University, New York, New York
- Morgan, Mr. E. Robins, Director of Placement, Lehigh University, Bethlehem, Pennsylvania
- Moritz, Mr. R. D., Director, Department of Educational Service, University of Nebraska, Lincoln, Nebraska
- Mueller, Dr. Kate Hevner, Dean of Women, Indiana University, Bloomington, Indiana
- Nichols, Dean Joseph C., Fenn College, 1983 E. 24th Street, Cleveland 15, Ohio
- Nield, Mrs. Majory P., Director, Vocational Office, Smith College, Northampton, Massachusetts
- Nielsen, Mr. Otto Richard, Texas College of Arts and Industries, Kingsville, Texas
- Northrup, Miss Catherine, Assistant Director, Student Personnel, Colorado State College, Fort Collins, Colorado
- Odum, Dr. Charles L., Assistant Professor of Psychology, Southwestern Louisiana Institute, Lafayette, Louisiana
- Olson, Miss Dorothy, Director, Texas Union, University of Texas, Austin 12, Texas
- Onthank, Dr. Karl W., Dean of Personnel Administration, Johnson Hall, University of Oregon, Eugene, Oregon
- Orr, Dean Cora L., Muskingum College, New Concord, Ohio
- Palmerton, Mr. L. R., Director Student Personnel, South Dakota School of Mines, Rapid City, South Dakota
- Parks, Mr. Donald S., Personnel Director, University of Toledo, Toledo, Ohio
- Partch, Dean C. E., School of Education, Rutgers University, New Brunswick, New Jersey
- Pattee, Mr. Howard H., General Secretary, California Association of Independent Secondary Schools, 645 W. 10th Street, Claremont, California
- Paul, Dr. J. B., Director, Bureau of Research, Iowa State Teachers College, Cedar Falls, Iowa
- Pearl, Miss Penelope M., Psychological Clinic, University of Michigan, 1027 E. Huron, Ann Arbor, Michigan
- Peck, Miss Margaret, Assistant Dean of Women and Counselor of Residence, University Residences for Women, University of Texas Austin 12, Texas
- Penoi, Mrs. C. R., Faculty Exchange, University of Oklahoma, Norman, Oklahoma
- Pepinsky, Prof. Harold B., Assistant Director and Assistant Professor of Education, University of Kansas, Guidance Bureau, Lawrence, Kansas
- Perry, Mr. William D., University of North Carolina, Chapel Hill, North Carolina
- Plank, Prof. William B., Department of Mining and Metallurgical Engineering, Lafayette College, Easton, Pennsylvania
- Porter, Prof. J. P., Department of Psychology, Box 87, Ohio University, Athens, Ohio
- Prince, Mr. Chester J., Personnel Director, William Jewell College, Liberty, Missouri
- Purdum, Mr. T. Luther, Director, Bureau of Appointments and Occupational Information, University of Michigan, Ann Arbor, Michigan

- Ramsay, Miss L Alice, Director, Personnel Bureau, Connecticut College, New London, Connecticut
- Read, Miss Julia E, Personnel Director, College of St Elizabeth, Convent Station, New Jersey
- Reed, Dr Anna Y, Department of Education Library, Chicago University, Chicago, Illinois
- Reeves, Mrs Mildred Jones, 363 Park Street, Hackensack, New Jersey
- Reilly, Miss Margaret M, Dean, Barat College of Sacred Heart, Lake Forest, Illinois
- Reinecke, Miss Virginia J, Counselor of Women, Faculty Exchange, University of Oklahoma, Norman, Oklahoma
- Replogle, Dr Fred A, Consulting Psychologist, Stevenson, Jordon and Harrison, 205 W Wacker Drive, Chicago, Illinois
- Reusser, Dr Walter C, Director, Student Personnel and Guidance, University of Wyoming, Laramie, Wyoming
- Richardson, Mr Thomas F, Director, Student Personnel, Texas Christian University, Fort Worth, Texas
- Riley, Miss Marguerite A, Placement Director, St Joseph College, Asylum Avenue, West Hartford, Connecticut
- Robison, Miss Sallie E, Personnel Director, Louisiana Polytechnic Institute, Ruston, Louisiana
- Roy, Mr Howard L, 507 Oak Street, S.E, University of Minnesota, Minneapolis, Minnesota
- Ruedisili, Dean C H., 103 South Hall, University of Wisconsin, Madison 6, Wisconsin
- Ruhlman, Miss Jessie L, College of Home Economics, Cornell University, Ithaca, New York
- Russell, Mr John Dale, Dean of Students, Division of Social Science, University of Chicago, Chicago, Illinois
- Sackett, Dr Everett B, Registrar, Director of Admissions, University of New Hampshire, Durham, New Hampshire
- Sage, Mr Nathaniel M, Placement Officer, Massachusetts Institute of Technology, Cambridge, Massachusetts
- Saul, Mr Herbert W, 18 Talbot Road, Braintree, Massachusetts
- Schaper, Dr Florence W, Director, Student Guidance and Personnel, Lindenwood College, St Charles, Missouri
- Schneiders, Dr Alexander A, Director of Student Personnel, Loyola University, Chicago, Illinois
- Schneider, Guendolen G, Vocational Advisement Supervisor, Veterans Administration, 17th and H Streets, N.W, Washington 25, D. C.
- Schonfeld, Dr Louise Rost, 211 W 106th Street, New York 25, New York
- Seaman, Mr William H., Director Admissions and Bureau of Appointments, Oberlin College, Oberlin, Ohio
- Seamans, Mr Harry W, 2810 Naylor Road, S.E, Apt. A137, Washington, D. C.
- Seibert, Mr Earl W, Director of Personnel, Green Mountain Junior College, Poultney, Vermont
- Shank, Mr Donald J, Administrative Associate, American Council on Education, 744 Jackson Place, Washington 6, D. C.
- Shell, Miss Helen H, Wright Aeronautical Co, Marlboro Inn, Montclair, New Jersey
- Sisson, Lt Donald E, Classification and Replacement, 270 Madison Avenue, New York, New York
- Sistrunk, Mr. Henry L, Y. M. C. A., 19 S. LaSalle Street, Chicago, Illinois
- Skinner, Dr. Laila, Dean of Women, Allegheny College, Meadville, Pennsylvania
- Smiley, Mr. E. Kenneth, Director of Admissions, Lehigh University, Bethlehem, Pennsylvania
- Smith, Sgt Harley B, 37354849, Psychological Research Project, Nav. Sec. E, Ellington Field, Texas
- Smith, Dr Margaret Ruth, Director of Student Activities, Wayne University, Detroit, Michigan
- Smoke, Mr Kenneth L, Professor of Psychology, Juniata College, Huntingdon, Pennsylvania

- Smutz, Mr Harold T , Regional Personnel Utilization Consultant, 635 New Federal Building, St Louis 1, Missouri
- Spindt, Mr Herman, Bureau of Guidance and Placement, University of California, Berkeley, California
- St Clair, Mr Walter F., 25 Waldo Street, Manchester, New Hampshire
- Stapleton, Miss Mary R., Associate in Personnel Bureau, Brooklyn College, Bedford Avenue and Avenue H, Brooklyn 10, New York
- Stedman, Miss Edith G., Director, Appointment Bureau, Radcliffe College, Cambridge, Massachusetts
- Stephanus, Miss Jane E., Acting Placement Director, University of Detroit, McNichols Road, Detroit 21, Michigan
- Stevens, Prof Edwin B., College of Education, University of Washington, Seattle, Washington
- Stogdill, Dr Emily L., Department of Psychology, Ohio State University, Columbus, Ohio
- Spacie, Mr Edwin G., Student Counselor, Michigan College of Mining and Technology, Houghton, Michigan
- Strang, Prof Ruth, Records and Research, College of Education, Teachers College, Columbia University, New York, New York
- Stromberg, Mrs Jean T., Assistant to the Counselor, One Sage Avenue, Cornell University, Ithaca, New York
- Stromberg, Lt. E L., U S. Naval Air Station, New Orleans, Louisiana
- Stuit, Lt (jg), Dewey B., USNR, Assistant Officer in Charge of Test and Research Section, Bureau of Naval Personnel, Washington, D C.
- Super, Capt Donald E., Psychological Branch, Unit T, AAFR and CH, Miami Beach, Florida
- Sutherland, Mr. Robert L., University of Texas, Austin 12, Texas
- Swann, Lt Reginald L, H (S) USNR, U.S.N.A.A.S., Rodd Field, Corpus Christi, Texas
- Swanson, Prof. Donald E., Director of Student Personnel, Hamline University, St. Paul, Minnesota
- Swenson, Miss Hilda G., Social Director, Box 151, College Station, Washington State College, Pullman, Washington
- Tallmadge, Miss Frances M., 328 W. 88th Street, New York 24, New York
- Tarbox, Mr Sidney, Director of Recruitment of 7th Regiment of U S Civil Service Commission, 12 S Monroe Street, Hinsdale, Illinois
- Tate, Mr. William, Dean of Students, University of Georgia, Athens, Georgia
- Taylor, Mary E., Women's Residence Center, University of Indiana, Bloomington, Indiana
- Thisted, Mr. M. N., Dean of Men, Western Illinois State Teachers College, Macomb, Illinois
- Thompson, Mr C Woody, Director, Office of Student Affairs, University of Iowa, Iowa City, Iowa
- Thornbury, Miss Zita L., Director, Vocational Bureau, Vassar College, Poughkeepsie, New York
- Thorpe, Miss Alice L., Appointment Office, Wheaton College, Norton, Massachusetts
- Tibbals, Dean C. A., Illinois Institute of Technology, Chicago, Illinois
- Tibbetts, Mr. Clark, Director of Institute of Human Adjustment, University of Michigan, Ann Arbor, Michigan
- Tiner, President Hugh M., George Pepperdine College, 1121 W 79th Street, Los Angeles, California
- Todd, Mr. J. Edward, 822 E John Street, Appleton, Wisconsin
- Trabue, Dean M R., School of Education, Pennsylvania State College, State College, Pennsylvania
- Triggs, Dr Frances O., American Nurses Association, 1790 Broadway, New York, New York
- Turner, Mrs Lillian Barbour, Shurtleff College, Alton, Illinois
- Twyman, Mrs Margaret G., Director, Placement Service, University of Oklahoma, Norman, Oklahoma
- Tyler, Miss Leona E., Director of Bureau of Personnel Research, University of Oregon, Eugene, Oregon

- Tyson, Lt Robert, 448 Riverside Drive, New York, New York
- Van Dusen, Ensign Albert C, H-V(S), USNR, Aerial Free Gunnery, Central Standardization Committee, NAOTC, Naval Air Station, Jacksonville, Florida
- Van Wagenen, Dr. (Miss) Beulah C, Girl Scouts, Inc, 155 East 44th Street, New York, New York
- Voorhees, Dean Edwin C, University of California, Berkeley, California
- Voorhees, Miss Helen M, Director, Appointment Bureau, Mount Holyoke College, South Hadley, Massachusetts
- Wallace, Mrs Frances Gibson, Freshman Adviser, 1520 5th Avenue, Huntingdon, West Virginia
- Wallace, Dr. Isabel King, Vocational Counselor for Women, University of Rochester, Rochester, New York
- Waldrop, Chaplain Robert S., U.S.S Benevolence, AH-13, c/o Fleet Post Office, New York, New York
- Walters, Dr J E, Personnel and Industrial Relations, McKinsey and Co, 60 E 42nd Street, New York, New York
- Wangeman, Dr Charles E, Head of Bureau of Placement, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
- Wegener, Miss Mary A, Appointments Office, Columbia University, New York, New York
- Weir, Miss Edith, Director, Bureau of Teacher Placement, University of Southern California, Los Angeles, California
- Westover, Mrs Ada S, Assistant to Dean of Women, University of Nebraska, 1900 Lyons Street, Lincoln, Nebraska
- Westover, Dr Frederick L, Student Personnel Assistant to the Dean, Brooklyn College, Bedford Avenue and Avenue H, Brooklyn, New York
- Whiteman, Mrs Harriet E, Educational Secretary, New Jersey State Teachers College, Newark, New Jersey
- Wightwick, Miss M Irene, Personnel Director, College of New Rochelle, New Rochelle, New York
- Wilcox, Prof E R, College of Engineering, University of Washington, Seattle, Washington
- William, Major Cornelia DeCamp, 2012 S 5th Street, Arlington, Virginia
- Williams, Prof Lewis W., Secretary, Appointment Committee, College of Education, University of Illinois, Urbana, Illinois
- Williamson, Prof E. G, Dean of Students, University of Minnesota, Minneapolis, Minnesota
- Wilson, Mr Eugene S., Amherst College, Amherst, Massachusetts
- Winegar, Mrs R M., 1303 Milton Way, San Jose, California
- Wunningham, Miss Jane Louise, Huron Valley Children's Center, 104 Welch Hall, Ypsilanti, Michigan
- Woellner, Prof Robert C, Secretary, Board of Vocational Guidance and Placement, University of Chicago, Chicago, Illinois
- Wood, Mr Austin B, Assistant Professor of Psychology, Brooklyn College, Brooklyn, New York
- Woodhouse, Mrs Chase Going, Director, Institute of Women's Professional Relations, Connecticut College, New London, Connecticut
- Wrenn, Dr C Gilbert, 755 North Church Street, Salem, Oregon
- Zerfoss, Prof. K. P, Department of Psychology, George Williams College, 5315 Drexell Avenue, Chicago, Illinois

MEASUREMENT ABSTRACTS*

Achard, F H and Clarke, Florence H "You Can Measure the Probability of Success as a Supervisor" *Personnel*, XXI (1945), 353-373

The authors studied 300 successful and indifferent supervisors in a public utility. They developed a supervisory rating scale which they found to be quite accurate and reliable for the four groups of supervisors studied. Successful supervisors had higher scores in tests of mental ability, personality and breadth of interest, and in ability to see and understand quickly (visual perception). The procedure proposed by the authors for the selection of supervisors includes: the selection of tests, the selection of candidates, sifting out of candidates on the basis of test scores, and final selection by executives, adding to ratings based on test scores (a) job knowledge, (b) versatility and adaptability, (c) mental, physical, and emotional fitness, (d) standing with colleagues, and (e) how well they "click" with prospective superiors. *Elizabeth Bell*

Allen, Mildred. "Relationship Between the Indices of Intelligence Derived from the Kuhlmann-Anderson Intelligence Tests for Grade I and the Same Tests for Grade IV" *Journal of Educational Psychology*, XXXVI (1945), 252-256

The Kuhlmann-Anderson Intelligence Tests for Grade I were administered to 300 school children in the middle of their first year. Near the beginning of their fourth grade they were given the same test for that level. Neither M A, I.Q., nor P c A v obtained at Grade IV were adequately predicted from the same scores in Grade I. The author believes the low relationship might be due to the differences in verbal content between the two tests (the first grade is all non-verbal) and that the results question the validity of long-range predictions from a group intelligence test at the non-verbal grade level. She concluded that a verbal group intelligence test has greater validity for predicting successful achievement in the tool subjects when given with an achievement test. This would also assist "in determining whether pupils are working up to their ability" *Elizabeth Bell*

Altus, William D. "Racial and Bi-Lingual Group Differences in Predictability and in Mean Aptitude Test Scores in an Army Special Training Center." *Psychological Bulletin*, XLII (1945), 310-320.

The dichotomous disposition of four bi-lingual groups, American Indian, Mexican, Filipino, and Chinese, permitted validating testing devices by the bi-serial correlation method. The highest average bi-serial correlations between four verbal subtests of the *Wechsler Mental Ability Scale*, Form B, and discharge versus graduation, were found in the Negro group, including three in the .50's. The Indian was next, and the Mexican lowest. The Filipino would have been second in average validity but for a low arithmetic subtest correlation. Verbal tests gave lower scores to the bi-lingual groups, but did not lose in validity. In mean test scores whites were definitely superior, Negroes were second, and Mexicans third. There was little difference between Chinese and Filipino, while the Indian was significantly lowest. The data prove nothing as to innate racial differences because of divergent backgrounds. *Elizabeth Bell*

Glick, H N, Flynn, Elizabeth, and Macomber, Lois "Some Comparisons Between the Original and the Revised Stanford-Binet Scales." *Journal of Educational Psychology*, XXXVI (1945), 177-183.

This article presents results of two studies conducted at Massachusetts State College with respect to certain comparisons between the original and revised Stan-

* Edited by Forrest A Kingsbury

ford-Binet scales. The most significant finding was that the two scales measured practically the same thing. The new scale was found to be more valid as a measure of intelligence in testing slightly superior children. Also it was found that the new scale tested slightly higher and showed greater dispersion of I Q's than the old. *Jane Nathier*

Goldfarb, William. "Note on a Revised Block Design Test as a Measure of Abstract Performance." *Journal of Educational Psychology*, XXXVI (1945), 247-251.

The *Revised Block Design Test* is an adaptation of the *Wechsler Block Design Test*. The adaptation involved two changes. The first was an extension of the time limit to five minutes for each card. The second was the introduction of a new scoring method whereby the subject received anywhere from zero to three credits per picture depending on the degree of accuracy. The test, consisting of seven designs, was administered to thirty adolescents with a mean age of 12.3. Each child was rated on the *Revised Block Design Test*, the *Wechsler Block Design Test*, the *Wechsler Similarities Test*, the *Vigotsky Test*, and the *Weigl Test*. All correlations were significant at the one per cent level. The correlation between the block design tests was .90. Both block tests offer a good measure of the ability to generalize. The *Revised Block Design Test* has a slightly higher correlation with other criteria of abstract ability. *Betty Steele*

Lewis, W. Drayton. "The Relative Intellectual Achievement of Mentally Gifted and Retarded Children." *Journal of Experimental Education*, XIII (1944), 98-109.

The purpose of the investigation was to determine the degree of relative achievement of mentally retarded and mentally gifted children. The subjects consisted of children from four hundred and fifty-five schools from thirty-six states. The *Kuhlmann-Anderson* and the *Unit Scales of Attainment Tests* were given to twelve hundred subjects for each grade. Reading, geography, arithmetic, and language tests were selected from a battery of eleven tests. The basic assumption was that if either group were working up to capacity, the median achievement scores should diverge as much from the norms as do the median mental ages. The results show that the superior group fails to achieve expectations, whereas the retarded group achieves more than expectations. The author concludes that relative achievement, based solely on mental age placements, is not significant. The important factors may be chronological age and length of attendance in school. *Betty Steele*.

London, Ivan D. "Psychology and Heisenberg's Principle of Indeterminacy." *Psychological Review*, LII (1945), 162-168.

The purpose of this article is to demonstrate that Heisenberg's principle of indeterminacy is inapplicable for psychology. Statisticians in psychology are not justified in presenting this principle as a sufficient reason for the inability to eliminate or evaluate interference. The principle results from the attempt to extrapolate into the microcosm the same dual concepts of space and time attributed to macrocosmic events. The principle applies only to the sub-atomic world. Interferences in psychological situations are the results of methodological difficulties inherent in the situation and not because of Heisenberg's principle. There are two types of phenomena in psychology. (1) Convergent phenomena in which behavior is determined from the average behavior of its parts; (2) Divergent phenomena in which a single discontinuous event affects the whole aggregate. In the latter phenomena the indeterminacy deriving from the unpredictable behavior of a single electron deals with isolated and unique electrons which cannot be duplicated. Such indeterminacy cannot be made the basis of a systematic psychology. *Betty Steele*.

Montagu, M. F. Ashley. "Intelligence of Northern Negroes and Southern Whites in the First World War." *American Journal of Psychology*, LVIII (1945), 161-188.

Scores made on the intelligence tests administered by the U. S. Army to Negro and white draftees during World War I are analyzed and discussed. Whites in any given state scored higher than Negroes in the same state on all tests and in all states.

except Kentucky and Ohio, where the Negroes excelled on the *Beta* tests. However, it was also found that the median score made by Negroes in a number of states was better than that made by whites in several other states. The conclusion, supported by geographical evidence, is drawn that the lower scores of the whites, in the instances where they appear, are due to inferior socio-economic conditions, and the generally lower scores of the Negroes are similarly explained. Findings indicate that differences in performance on the tests between Negroes and whites are due solely to the action upon native endowment of differences in socio-economic history. *Frances E. Smith*

Munroe, Ruth. "Three Diagnostic Methods Applied to Sally." *Journal of Abnormal and Social Psychology*, XL (1945), 215-227.

Sally is the fictitious name applied to a normal, fairly well adjusted, Sarah Lawrence College-student, one of a group of eleven who were selected for an intensive study to determine with what accuracy certain testing procedures could portray the wholeness of personality and educative potentialities of the students. The *Rorschach* test, modified for large-scale use, was administered by Munroe, a graphological analysis was effected by Lewinson, and Waehner made an appraisal from spontaneous drawings. These findings were augmented by reports from the faculty adviser and instructors. The experimenters and teachers worked blindly and independently of each other. There was a sufficient degree of conformity between the reports, the analysis of the tests, and the accomplishments of the students to suggest that remedial methods indicated by the tests might have been of value. *Helen Heath*

Patterson, R. Melcher. "Evaluation of a Prolonged Pre-Academic Program for High Grade Mentally Deficient Children in Terms of Subsequent Progress." *Journal of Experimental Education*, XIII (1944), 86-89.

The Prolonged Pre-Academic Program is an experimental educational unit at the Wayne County Training School. On the basis of experimental data it was found that mentally deficient children of I.Q. levels 60-79 inclusive, educated on the time schedule described, appeared to progress in academic work when it was introduced at a rate which enabled them to arrive at approximately the same average grade level by 16 years of age as they would have achieved had academic instruction started at the earliest possible mental age and been given continuously. *Jane Nathler*

Rabin, Albert I. "Psychometric Trends in Senility and Psychoses of the Senium." *Journal of General Psychology*, XXXII (1945), 149-162.

The *Wechsler-Bellevue* intelligence scale was administered to 150 individuals aged 60-84, admitted to the New Hampshire State Hospital, and 100 of the resulting records were divided into 4 diagnostic groups, Senile, C.A.S., Miscellaneous (not including organic), and Non-Psychotic. Analysis of results showed no intra-test organizational or structural differences between classes. Large positive deviations occurred in the 3 major verbal subtests, while practically all performance subtests showed negative deviations. When scores of each subject were correlated with age, only performance subtests showed statistically significant negative correlations, though *r*'s for verbal subtests were also negative. Changes occurring in performance and timed subtests appear to be due to age levels rather than differential diagnosis, with inflexibility and perseverative tendency as the significant factors in lowered performance. Findings of this investigation fail to agree with Wechsler's dichotomy of tests which "do not hold up with age," and do not support the statement that schizophrenia resembles "premature aging." *Frances E. Smith*

Ryan, Thomas Arthur. "Merit Rating Criticized." *Personnel Journal*, XXIV (1945), 6-15.

The author suggests the use of his *Inventory of Personnel* as a simpler procedure based on "few doubtful assumptions" than the merit rating, until new and different methods for evaluating personnel are produced through research. He criticizes the graphic rating scale of today as being based on false logic, since it secures ratings on distinct traits, then assigns a total score ("adding apples and carts"), thus substituting one trait for another. Also, there is no known way of

validating the weights given traits. The halo effect he believes so strong that a single careful rating of a man's "over-all value" is probably better. Ratings cannot be compared or corrected for constant errors with confidence. He also states that studies have shown their reliability to be generally low. *Elizabeth Bell*

Sargent, Helen. "Projective Methods: Their Origins, Theory, and Application in Personality Research." *Psychological Bulletin*, XLII (1945), 257-293.

The theoretical and historical background of the projective methods is presented, and the methods are discussed with regard to their application, under the classifications of materials, functional uses, techniques, and purposes. Fundamental experiments testing the efficacy of the methods have been few, and there is need for further research bearing on underlying assumptions and predictive value. Experiments thus far conducted demonstrate the basic mechanism of projection and claim some degree of predictive success for projective methods. Methodological problems include questions of reliability, validity, and standardization. Evaluation of the methods stresses the importance of an open-minded attitude toward theory revision and innovations in method. Projective methods hold promise in clinical psychology for a science of diagnosis and treatment, there is considerable evidence that they are worth thorough exploration. The bibliography contains 274 references. *Frances E. Smith*

Smith, Francis F. "The Use of Previous Record in Estimating College Success." *Journal of Educational Psychology*, XXXVI (1945), 167-175.

The purpose of the investigation was to predict, on the basis of high-school records, the number of grade-point scores each freshman would earn in his first semester in college. The high-school record, based on about eighty examinations evaluated by sixteen instructors at different times, was converted into one figure indicating the percentage of recommended grades. A multiple-prediction formula was used which included the freshman's aptitude percentile, reading percentile, and English A percentile. The subjects were 903 freshman students. The standard error of estimate from the multiple-prediction formula was 7.97. The results show a correlation of .60 between the high-school record based on a single figure and actual success in the first semester in college. However, the score loses its predictive value after a year or more. It was found that the best indicator of future scholastic success was the previous semester's record. *Betty Steele*.

Swift, Joan W. "Reliability of Rorschach Scoring Categories with Pre-School Children." *Child Development*, XV (1944), 207-216.

Reliability for Rorschach scoring categories was tested with pre-school children as subjects under the following conditions: (1) Test-retest over a 30-day interval. Reliabilities ranged from +.15 to +.83. (2) Test-retest over a 14-day interval with a second series of ink blots interpolated on the 7th day. Uncorrected r 's ranged from +.41 to +.74, and corrected values for scores obtained by combining test and retest from +.59 to +.84. (3) Use of a parallel series of ink blots over a 7-day interval, with subjects identical with those used in (2). The r 's ranged from -.06 to +.84, all but one being above +.39, while corrected values for scores obtained by combining the 2 series of cards ranged from -.08 to +.86. (4) Test-retest over a ten-month interval, with 20 subjects retested as part of (2) ten months after participation in (1). Range of r 's was from +.18 to +.53. Reliability was affected by small number of responses given by each child, low frequency of occurrence of many response categories, and variability of interest and attention span of the children. *Frances E. Smith*.

Swift, Joan W. "Matchings of Teachers' Description and Rorschach Analyses of Pre-School Children." *Child Development*, XV (1944), 217-224.

Investigation was made of the validity of the total personality picture given by the Rorschach method when applied to pre-school children. Validating material consisted of a 250-word personality sketch of each of 30 pre-school children, written by the children's teachers. Of the thirty matchings, 14 were correct and 16 incorrect, a result found to be significantly better than chance at the 1 per cent level of confi-

dence. All but 3 of the 14 correct matchings were of boys, a result possibly due to sex differences in stereotypy of behavior *Frances E Smith*

Wells, F L "Mental Factors in Adjustment to Higher Education" *Journal of Consulting Psychology*, IX (1945), 267-286.

"An introduction to the psychometric work of the Grant Study, concerned with well-adjusted Harvard undergraduates," with reference to the "technical problems in the measurement of the superior intellect, career choice, as well as educational prognosis" It is concluded that the abstraction of the liberal arts college curriculum is not the real problem, but the fact that it is offered to many who probably cannot assimilate it "A more highly ideational type of mind than commonly receives it" profits most from it One half of the present college population could be replaced from the ranks of those economically not so fortunate Two broad mental qualities are needed for college scholastic success, absorptiveness and creativity The latter is hard to recognize with our present methods of measurement For this reason, objective tests are believed inadequate for higher education Eight case studies of well-adjusted students are presented *Elizabeth Bell*

Woodrow, Herbert "Intelligence and Improvement in School Subjects." *Journal of Educational Psychology*, XXXVI (1945), 155-166.

Studies were conducted at Litchfield, Ill. and Providence, R I, to determine whether improvement is synonymous with intelligence and whether there is any common factor in improvement In the Litchfield study, the *Metropolitan Achievement Battery* and the *Otis Quick-Scoring Mental Ability Tests* were given to fourth-, fifth-, and sixth-grade children in May of successive years Intercorrelations of gains in six school subjects were 12 Factor analyses revealed no gain factor common to the six subjects Three factors were found, 1) an intelligence factor determining the IQ, but not the gain, 2) an arithmetic factor, 3) a reading and English factor In the Providence study the *Stanford Achievement Test* was used The differences in standard scores for periods of one, two, and three years revealed low gain correlations Three factors were found which were similar to those found in the Litchfield study The conclusion of both studies is that there is no general gain factor, and that there is no significant relation between change in score and IQ *Bette Steele*

EDUCATIONAL and
PSYCHOLOGICAL



MEASUREMENT

VOLUME FIVE, NUMBER FOUR, WINTER

<i>Evidence on the Validity of the Armed Forces Institute Tests of General Educational Development (College Level).</i>	
HENRY S. DYER	321
<i>A Study of the Factor Structure of Thirteen Personality Variables.</i>	
CONSTANCE LOVELL	335
<i>The Office of Radio Research: A Division of the Bureau of Applied Social Research, Columbia University.</i>	
MARJORIE FISKE and PAUL F. LAZARSFELD	351
<i>The Duties of Civil Service Examiners and Test Technicians</i>	371
<i>The Arrangement of Choices in Multiple Choice Questions and a Scheme for Randomizing Choices.</i>	
CHARLES I. MOSIER and HELEN G. PRICE	379
<i>The Rationale of Temperament Testing.</i>	
DONCASTER G. HUMM	383
<i>Mechanical Ability, Its Nature and Measurement. II.</i>	
<i>Manual Dexterity.</i>	
J. R. WITTENBORN	395
<i>Speed and Level Components in Time-Limit Scores: A Factor Analysis.</i>	
WILLIAM M. DAVIDSON and JOHN B. CARROLL	411
<i>The Use of an Objective Test in Predicting Rhetoric Grades.</i>	
IRWIN A. BERG, GRAHAM JOHNSON, and ROBERT P. LARSEN	429
<i>A Quick Graphic Method for Product Moment 'r.'</i>	
WILLIAM LEROY JENKINS	437
<i>Measurement News</i>	445
<i>The Contributors</i>	449
<i>Measurement Abstracts</i>	452

The Contents of This Issue Are Listed in the *Education Index*

CIRCULATION, ETC.
24, 1912, AND MARCH
SUREMENT, published
CITY OF WASHINGTON,

for October 1, 1945

DISTRICT OF COLUMBIA

Before
appeared G
and says t
MEASUREMENT, and that the following is, to the best of his knowledge and belief,
a true statement of the ownership, management (and if a daily paper, the circula-
tion), etc., of the aforesaid publication for the date shown in the above caption,
required by the Act of August 24, 1912, as amended by the Act of March 3, 1933,
embodied in section 537, Postal Laws and Regulations, printed on the reverse of this
form, to wit:

1. That the names and addresses of the publisher, editor, managing editor, and
business managers are: Publisher, G. Frederic Kuder, 917 Fifteenth St., N.W., Wash-
ington, D. C. Editor, G. Frederic Kuder, 917 Fifteenth St., N.W., Washington, D. C.
Managing Editor, G. Frederic Kuder, 917 Fifteenth St., N.W., Washington, D. C.
Business Manager, G. Frederic Kuder, 917 Fifteenth St., N.W., Washington, D. C.

2. That the owner is: (If owned by a corporation, its name and address must be
stated and also immediately thereunder the names and addresses of stockholders own-
ing or holding one per cent or more of total amount of stock. If not owned by a
corporation, the names and addresses of the individual owners must be given. If
owned by a firm, company, or other unincorporated concern, its name and address, as
well as those of each individual member, must be given.) G. Frederic Kuder, 917
Fifteenth Street, N.W., Washington, D. C.

3. That the known bondholders, mortgagees, and other security holders owning
or holding 1 per cent or more of total amount of bonds, mortgages, or other securities
are: (If there are none, so state) None.

4. That the two paragraphs next above, giving the names of the owners, stock-
holders, and security holders, if any, contain not only the list of stockholders and
security holders as they appear upon the books of the company but also, in cases
where the stockholder or security holder appears upon the books of the company as
trustee or in any other fiduciary relation, the name of the person or corporation for
whom such trustee is acting, is given, also that the said two paragraphs contain
statements embracing affiant's full knowledge and belief as to the circumstances and
conditions under which stockholders and security holders who do not appear upon the
books of the company as trustees, hold stock and securities in a capacity other than
that of a bona fide owner; and this affiant has no reason to believe that any other
person, association, or corporation has any interest direct or indirect in the said
stock, bonds, or other securities than as so stated by him.

5. That the average number of copies of each issue of this publication sold or
distributed, through the mails or otherwise, to paid subscribers during the twelve
months preceding the date shown above is: not a daily. (This information is required
from daily publications only.)

Signed: G. Frederic Kuder, Editor. Sworn to and subscribed before me this
28th day of September, 1945—George L. Haines
(Seal)

(My commission expires January 31, 1949.)

EVIDENCE ON THE VALIDITY OF THE ARMED FORCES INSTITUTE TESTS OF GENERAL EDUCATIONAL DEVELOPMENT (COLLEGE LEVEL)

HENRY S. DYER
Harvard University

Purpose of the Study

THE nature and purposes of the *United States Armed Forces Institute Tests of General Educational Development* are fully described in the *Examiner's Manual* provided with the tests.¹ The battery consists of four tests as follows:

- Test 1. *Correctness and Effectiveness of Expression.*
- Test 2. *Interpretation of Reading Materials in the Social Studies.*
- Test 3. *Interpretation of Reading Materials in the Natural Sciences.*
- Test 4. *Interpretation of Literary Materials.*

All four of the tests are objective in form, the questions being wholly of the multiple-choice type. There are two equivalent forms of each test, one of which is to be administered exclusively by the Armed Forces Institute (the Military Form) and the other of which is available to colleges generally through the American Council on Education (the Civilian Form).

According to the *Examiner's Manual*:

"The college level tests are intended for use primarily to determine whether or not the individual tested is as capable of carrying on advanced college work as the student who has taken certain broad introductory or survey courses generally offered in the first two years of the liberal arts college, or has

¹ U. S. Armed Forces Institute, *Tests of General Educational Development (College Level)*. *Examiner's Manual*. New York. American Council on Education, 1944.

reached the same level of general educational development as the student who has had such survey courses."²

The present study was undertaken to discover whether the results of the tests were sufficiently valid for use with veterans who might seek admission to Harvard after the war. Specifically the answers to three questions have been sought:

1. Do the test results provide a basis for placing students in advanced standing at Harvard?
2. Do they provide a sound basis for the selection of candidates for admission to Harvard?
3. Can they be used in counseling the veteran on his choice of a field of concentration?

Crawford and Burnham³ made a study of Yale freshmen which would lead one to expect an affirmative answer to the second question. They found that the total of the standard scores on the A.F.I. Tests "correlated as well with Freshman first term averages in all courses, as did the average of all College Board Achievement Tests."⁴

Limitations of the Present Study

With veterans actually seeking admission to colleges in increasing numbers, it is not possible to wait for the appropriate amount and kind of data to accumulate on the A.F.I. Tests before deciding to use them or not to use them. The data of the present study provide no final answers to the questions proposed, but it is hoped they may furnish some helpful clues useful to college administrative officers.

The group studied at Harvard was composed of undergraduates whose educational careers, for the most part, had not been interrupted by military service. Their performance on the tests cannot, therefore, be considered as directly comparable to that of the college-minded veteran who not only will

² *Op. cit.*, p. 3.

³ A. B. Crawford and P. S. Burnham. "Trial at Yale University of the Armed Forces Institute General Educational Development Tests" *Educational and Psychological Measurement*, IV (1945), 261-270.

⁴ This comparison would have been more enlightening if the College Board's *Scholastic Aptitude Test* scores had been averaged in with the *Achievement Test* scores

have been away from formal classroom work for some time, but who will also have undergone experiences whose effect on his learning habits is, at best, difficult to predict. Furthermore, the Harvard group took the tests on a voluntary basis, motivated solely by patriotic considerations and the hope of receiving one of a series of monetary prizes. Under these conditions it was not expected that the group would constitute a representative sample even of a normal civilian undergraduate population. Its incentives could hardly be considered similar to those of the returning veterans.

There were 114 undergraduates who completed all four tests (civilian forms) and on whom there was the essential accessory information. The composition of this group is shown in Table 1.

TABLE 1
*Distribution of the Tested Group According to Class Standing and Fields of Concentration**

	Non-Scientific Fields	Scientific Fields	Totals
Freshmen	29	30	59
Sophomores	7	10	17
Juniors	15	12	27
Seniors	6	5	11
Totals	57	57	114

* Since the field of concentration is not formally elected until the beginning of the sophomore year, the freshmen were assigned to "probable fields" on the basis of expressed preference.

Two indices were available by which the general scholastic ability of the tested group could be compared to that of a normal prewar class. The first of these was the Verbal Score on the *College Entrance Board Scholastic Aptitude Test* which is taken by nearly all students as part of the entrance examination. The second was the college rank at the end of the freshman year. The college rank is reported in seven groups: Group 1 represents a straight A record, and Group 7 represents an unsatisfactory record. Table 2 shows how the tested group compares with the Class of 1942 on these two indices. The tested group is relatively overweighted with students whose interests lie along scientific lines, but this fact should not seriously impair the results of the study if each of the groups is considered separately. There is, however, little question that

the tested group on the whole is sufficiently above average in scholastic ability to require that an allowance for the difference must be made in any general application of the findings. The allowance can be made with some confidence because of the fact that the range of ability in the tested group, as shown by the standard deviations, is not unlike that of the normal group. In other words, the tested group provides a reasonably good sampling of the less able as well as of the superior students.

There is one further important limitation on the present study. Any findings related to the usability of the A.F.I. Tests for placing students in advanced standing will probably not be generally applicable to colleges where the program of study and

TABLE 2
General Scholastic Ability of the Tested Group Compared with That of the Class of 1942

Area of Concentration	No. of Cases		C.E.E.B. Verbal Score				College Rank (Freshman Year)			
	Tested Group	Class of 1942	Tested Group		Class of 1942		Tested Group		Class of 1942	
			M	σ	M	σ	M	σ	M	σ
Non-Scientific	57	624	640.5	91.8	574.8	92.0	3.77	1.36	4.44	1.37
Scientific	57	265	637.2	103.6	551.3	90.0	3.95	1.76	4.18	1.45
Total	114	889	638.9	97.9	567.8	92.0	3.86	1.58	4.36	1.40

the system of promotion are unlike those at Harvard. The students in this study have not been exposed to so-called "survey" courses. Ordinarily, a student in Harvard College, by the time he completes his sophomore year, takes at least one course in the natural sciences, one in the social sciences, and one in the humanities. He is free to select any one of a large number of courses in each of these areas. There is thus no guarantee that he will have obtained a broad acquaintance with the material in any given area. The one exception to this rule is that practically all freshmen are required to take a course in English composition.

*Value of the A.F.I. Tests for Placement in
Advanced Standing*

In order to determine whether it is necessary in this study to differentiate between freshmen and upperclassmen, it has

seemed advisable to look first into the question of whether the A.F.I. Tests show any relationship to the number of terms of work that a student has completed in college. In other words, do the tests measure educational development as it is conceived at Harvard?

The representation from each class was so small and the difference in average scholastic ability among the several groups was so large that a simple comparison of mean scores from class to class would be all but meaningless. Therefore, in order to secure a reasonably adequate answer to the question of relationship between the test scores and the amount of academic work completed, the method of partial correlations has been employed. Although partial correlations in the present instance will not provide an absolutely rigorous statement of the situation, it is believed that they provide a practical approximation of the true picture.

We wish to know the correlations between the A.F.I. Test scores and number of college terms completed when general scholastic ability is held constant. For this purpose, it is essential that the measure of scholastic ability shall be based on evidence obtained before the student entered college. Such a measure is found in the Predicted Rank List Standing (PRL), an index computed routinely for every applicant to Harvard College. The PRL is a composite index based upon the applicant's secondary-school class rank (hereinafter called School Rank) and his scores on the College Entrance Board examinations. It normally has a correlation of about .65 with College Rank at the close of the freshman year. It is expressed in the same terms as the College Rank, that is, a PRL of 1 represents highly superior ability and a PRL of 7 represents inferior ability.

Table 3 shows the partial correlations between the various A.F.I. Test scores and the number of terms completed in college, with initial scholastic ability, as measured by the PRL, held constant.⁵

⁵ The School Rank is converted to a standard score with mean of 85 and standard deviation of 5. Complete tables of zero-order correlations will be found in Tables 8, 9, and 10 at the end of this article

Using Guilford's "Table of Significant Values of r , R , and t ,"⁶ we find that two of the partial correlations in Table 3 can be considered statistically significant. For the Scientific Group, the partial correlation of .28 between the *Social Studies Test* and terms completed is above the five per cent level of confidence and that between Total Score and terms completed (.24) is above the one per cent level. None of the partial correlations found for the Non-Scientific Group is statistically significant. In other words, the present data suggest that, to a small

TABLE 3

Partial Correlations between A.F.I. Test Scores and Number of Terms Completed in College, with Initial Ability Held Constant

	Non-Scientific Group (N=57) r	Scientific Group (N=57) r
Test 1 (Expression) . .	.03	.05
Test 2 (Social Studies)17	.28
Test 3 (Natural Science) ..	.12	.23
Test 4 (Literature)14	.01
Total Score*16	.24

* Total Score was obtained by summing the standard scores on the four tests.

extent, the A.F.I. Tests measure the educational development of students concentrating in science, but not of students concentrating in social studies and humanities. However, in view of the small magnitude of even the significant relationships, the findings on the present group indicate that the *A.F.I. Tests of General Educational Development* cannot be used as a basis for placing students in advanced standing in Harvard College unless a fundamental change were made in the principles governing promotion from one class to another. The value of the tests at colleges that offer specific courses in general education remains to be determined.

Value of the A.F.I. Tests for Selecting Students for Admission

In view of the small relationships found between the A.F.I. scores and the number of terms completed in college, it was felt

⁶ Guilford, J. D. *Psychometric Methods*. New York: McGraw-Hill Book Company, 1936. Pp 548-9.

that in studying the results further, the factor of number of terms completed could be disregarded. When selecting students for college study, one ordinarily tries to find the measures or combination of measures that are most predictive of academic success, where academic success is itself measured in some such terms as average freshman grades, grade-point averages, and the like. Mention has been made above of the Predicted Rank List (PRL) as the pre-admission index having the highest known correlation with College Rank at Harvard. Since the PRL is a composite of School Rank and College Entrance Board examinations scores, we shall combine the A.F.I. Test scores with School Rank in the same manner and

TABLE 4

*Correlations with College Rank at the Close of the Term in Which the A.F.I. Tests Were Given**

	Non-Scientific Group (N=57) r	Scientific Group (N=57) r	Total Group (N=114) r
PRL64	.71	.67
Total A.F.I. Score41	.52	.46
School Rank62	.63	.62
Total A.F.I. + School Rank† ..	.65	.66	.65

* Certain of the correlation coefficients are technically negative, but the negative signs have been omitted to avoid confusion in meaning.

† The values in this row are multiple R's. They are thus not strictly comparable to the r's obtained with the three other variables. That is, with a new sample one would expect shrinkage in the multiple R's beyond that to be expected in the zero-order r's.

compare the two composites on the basis of the degree to which they predict the College Ranks that were assigned at the close of the term in which the A.F.I. Tests were taken. Table 4 shows how the two composites compare with each other.

It is apparent from Table 4 that, for the tested group, the composite of the A.F.I. Total score and School Rank compares favorably with the PRL in the prediction of College Rank. It will be observed that the School Rank factor, which is common to both of the composites, accounts for a relatively large proportion of the predictable variance in both cases. One cannot say how far this factor will be affected by the interruption in education that the returning veteran will have experienced.

Its predictive power will probably vary from one applicant to another. However, it is scarcely a measure that one would wish to discard altogether, since there is no reason to suppose that the predictive power of the tests—both A.F.I. and College Board—will not also suffer in a similar fashion and for many of the same reasons.

Of some interest is the fact that while the PRL appears to give a slightly better prediction for the Scientific Group as compared to the Non-Scientific Group, the A.F.I.-School-Rank composite seems to predict the academic performance of both groups about equally well. The superiority of the PRL with science concentrators may well be due to the fact that the College Board examination includes a test of mathematical aptitude which is missing in the A.F.I. series. Crawford and Burnham in their study at Yale concluded that the College Board *Mathematical Aptitude Test* is "probably indispensable for scientific or engineering majors."⁷

In general, the evidence from this portion of the study seems to indicate that the A.F.I. Tests are useful as an aid in selecting students capable of college work at Harvard.

Value of the A.F.I. Tests as a Basis for Guidance

Do the *A.F.I. Tests of General Educational Development* provide the counselor with tools for advising the veteran with respect to his field of concentration? The data of the present study are inadequate to supply anything but the barest hint of an answer to this question.

It is probably not wholly unreasonable to assume that the Non-Scientific Group has, on the average, more ability in the fields of its choice than the Scientific Group, and that the Scientific Group has, on the average, more ability in scientific subjects than the Non-Scientific Group. It remains to be seen whether these expected differences in ability are matched by differences in performance on the A.F.I. Tests. Table 5 provides evidence on this point.

One finds that the Non-Scientific Group consistently tends to surpass the Scientific Group in the tests ordinarily associated

⁷ *Op. cit.*, p. 268

TABLE 5

Comparisons of Mean A.F.I. Scores Obtained by Two Groups of Concentrators

	Non-Scientific Group (N=57)			Scientific Group (N=57)			$M_1 - M_2$	Critical Ratio†	p
	M_1^*	σ_{M_1}	S.D. ₁	M_2^*	σ_{M_2}	S.D. ₂			
Test 1 (Written English)	98.8	1.26	9.4	96.7	1.24	9.3	+2.1	1.2	12
Test 2 (Social Studies) . .	72.2	1.47	11.0	68.5	1.48	11.1	+3.7	1.9	.03
Test 3 (Natural Science)	61.3	1.96	14.6	69.1	1.47	11.0	-7.8	3.2	.001 -
Test 4 (Literature)	66.8	1.34	10.0	63.5	1.29	9.6	+3.3	1.8	.04

* Raw scores were used in the computation of the means and standard deviations.

† The standard errors of the differences between the means were computed by means of the formula:

$$\sigma_{d_{12}} = \sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2}$$

with its field of interest, i.e., Written English, Social Studies, and Literature. With respect to the Social Studies and Literature tests, the differences between the two groups are statistically significant, that is, the likelihood is less than five per cent that differences of this size would arise as a matter of chance. Similarly, the difference between the mean scores on the Science test is in the direction one would expect, and this difference is of such size that one would expect it to occur as a matter of chance less than once in a thousand times.

The actual magnitude of these differences is, of course, not very large compared to the total range of the scores. However, if a student were to score relatively high on Tests 1, 2, and 4 and low on Test 3, one might at least hazard a guess that his abilities were more like those of students studying in the broad area of social studies and humanities than like those of students pursuing the sciences as a major field of interest.

As further evidence of the value of the A.F.I. Tests for guidance purposes, the degree of correlation of the tests with subsequent performance in actual fields of study is required. Unfortunately, the present data are not sufficiently numerous to provide unbiased criterion measures for each of the several fields. Nevertheless, it seems reasonable to suppose that the College Rank assigned to each student at the close of the term in which he took the A.F.I. Tests provides a rough measure of

TABLE 6
*Courses Taken by Two Groups of Concentrators in the Term When
 A.F.I. Tests Were Given*

	Non-Scientific Group (N=57)		Scientific Group (N=57)	
	No. of Courses	Average per Student	No. of Courses	Average per Student
Natural Science . .	36	6	136	2.4
Social Studies . . .	89	1.6	29	.5
Humanities	114	2.0	81	1.4
Total	239	4.2	246	4.3

performance in science for the Scientific Group and a similarly rough measure of performance in social studies and humanities for the Non-Scientific Group. The basis of this supposition is shown in Table 6.

On the average, the College Ranks for students in the Non-Scientific Group were computed on the basis of 4.2 courses of which 3.6 were taken in the social studies and humanities areas, and College Ranks for the Scientific Group were computed on the basis of 4.3 courses, of which 2.4 were taken in the natural science area. Clearly, the College Rank does not provide a "pure" criterion measure, but correlations based upon it may be considered approximately indicative of the predictive power of the several tests. One would expect to find that the Science

TABLE 7
Correlations of Each of the A.F.D. Tests with College Rank

	Non-Scientific Group (N=57)		Scientific Group (N=57)			Critical Ratio	p
	r_1^*	z_1	r_2^*	z_2	$z_1 - z_2$		
Test 1 (Written English) . .	.38	.40	.23	.23	.17	.9	.18
Test 2 (Social Studies)26	.27	.49	.54	-.27	1.4	.08
Test 3 (Natural Science) . .	.29	.30	.58	.66	-.36	1.9	.03
Test 4 (Literature)48	.52	.34	.35	.17	.9	.18

* The correlation coefficients are technically negative, but the signs have been dropped to avoid confusion in meaning. Each correlation has been converted to Fisher's z-value for the purpose of securing an exact test of the significance of the differences (See Fisher, R. A. *Statistical Methods for Research Workers*, New York, 1941 pp. 190 ff.) Since each of the z-values is based on the same number of cases, the standard error of z is a constant: .136. The standard error of the difference between each pair of z-values is also a constant: .192.

test has a higher correlation with the College Rank of the Scientific Group than with that of the Non-Scientific Group; and that the other three tests have a higher correlation with the College Rank of the Non-Scientific Group than with that of the Scientific Group. Table 7 gives the correlations for each group.

From Table 7, it is apparent that our expectations are borne out except in the case of the *Social Studies Test*. Here, for some reason, the Scientific Group correlation is higher than that of the Non-Scientific Group. It should be noted that this difference, large as it is, is nevertheless not statistically significant. One reason for the low correlation obtained on this test for the Non-Scientific Group may be that the ceiling on the test is not sufficiently high for students interested in social studies. In other words, the test may not be able to differentiate so well among persons who are well read in the field as among persons for whom social questions are, on the average, a secondary concern. The present findings also suggest that the A.F.I. *Social Studies Test* may be useful with the science concentrators as a predictor of general academic performance. The same cannot be said for students with a non-scientific turn of mind.

The one statistically significant difference in Table 7 is that between the two correlations involving the *Natural Science Test*. In this instance, the difference is in the direction expected, and from the size of the correlation coefficient for the Scientific Group, it seems fairly clear that the A.F.I. *Science Test* has a genuine value for guidance purposes.

As to the remaining two tests—*Written English* and the *Interpretation of Literary Materials*—the present study has produced no findings of any clear significance for prediction purposes. In a future study these two tests will be further investigated.

Summary of the Findings

The findings of this study are tentative and should be regarded with caution. The nature and size of the sample used makes impossible any definitive statement regarding the value and proper use of the A.F.I. *Tests of General Educational Development*. However, with specific reference to the group tested, the results of the study suggest the following:

TABLE 8
Intercorrelations, Means, Standard Deviations
Non-Scientific Group (N = 57)
A.F.I. Tests

	School Rank	PRL	Test1	Test2	Test3	Test4	Total	Terms Completed	College Rank
School Rank41	.26	.25	.36	.40	.28	.62
PRL61	.59	.58	.66	.73	.34	.64
Test 142	.58	.60	.74	.23	.38
Test 269	.68	.86	.33	.26
Test 364	.87	.29	.29
Test 486	.32	.48
Total35	.41
Terms Completed41
M	91.2	3.6	98.8	72.2	61.3	66.8	288.3	2.4	3.5
S.D.	3.2	.87	9.4	11.0	14.6	10.0	22.2	2.0	1.5

1. Although the A.F.I. Tests show a statistically significant relationship with the amount of work completed in Harvard College by science concentrators, the magnitude of the relationship is so small that the tests do not provide a sound basis for placing such students in advanced standing under the present system of promotion.

2. The A.F.I. Tests show no significant relationship with the amount of work completed by non-science concentrators in Harvard College. These two findings, however, should be interpreted in the light of the fact that the tested group was not exposed to a curriculum in "general education."

TABLE 9
Intercorrelations, Means, Standard Deviations
Scientific Group (N = 57)
A F I Tests

	School Rank	PRL	Test1	Test2	Test3	Test4	Total	Terms Completed	College Rank
School Rank39	.54	.54	.37	.59	.33	.63
PRL55	.69	.73	.53	.78	.30	.71
Test 149	.44	.43	.71	.21	.23
Test 279	.44	.89	.40	.49
Test 347	.83	.37	.58
Test 472	.15	.34
Total37	.52
Terms Completed29
M	91.3	3.5	96.7	68.5	69.1	63.5	286.3	2.1	3.9
S.D.	3.4	1.2	9.3	11.1	11.0	9.6	19.3	2.0	1.8

TABLE 10
Intercorrelations, Means, Standard Deviations
Total Group (N=114)
A F I Tests

	School Rank	PRL	Test1	Test2	Test3	Test4	Total	Terms Completed	Col-lege Rank
School Rank39	.40	.36	.36	.49	.30	.62
PRL56	.63	.61	.56	.74	.31	.67
Test 146	.46	.52	.73	.23	.31
Test 263	.58	.86	.37	.39
Test 349	.80	.28	.35
Test 479	.25	.42
Total36	.46
Terms Completed35
M	91.2	3.5	97.8	70.4	65.2	65.2	287.3	2.3	3.7
S.D.	3.3	1.0	9.4	11.2	13.5	10.0	20.8	1.0	1.7

3. The total score of the A.F.I. Tests when used in combination with the student's school rank should provide a reasonably good prediction of his subsequent academic success in College.

4. For the tested group, the A.F.I. Tests appear to measure "general aptitude" rather than "general educational development."

5. On the average, the score patterns yielded by the A.F.I. battery appear to differentiate slightly the students interested primarily in the social studies and humanities from those interested in the natural sciences.

6. The *A.F.I. Test in the Social Studies* may be useful with students of scientific bent as a predictor of general academic ability and development.

7. The *A.F.I. Test in the Natural Sciences* provides a useful instrument for predicting college success in the sciences.

A STUDY OF THE FACTOR STRUCTURE OF THIRTEEN PERSONALITY VARIABLES¹

CONSTANCE LOVELL
University of Southern California

Introduction

THE purpose of this study was to make a factor analysis of the thirteen variables of personality measured by Guilford's *Inventory of Factors STDCR*, the Guilford-Martin *Inventory of Factors GAMIN*, and the Guilford-Martin *Personnel Inventory I*.

These inventories were constructed to measure those personality characteristics which previous factor analysis and clinical work had indicated as important.² The original studies showed that the thirteen factors were not completely independent of each other though they were sufficiently separate to make individual scores helpful. The present study has involved factor analysis of the correlations found between them for the purpose of determining the clusters into which they fall. In other words, it has been designed to investigate the nature of more generalized super-factors with which the specific and interrelated original factors are loaded. Because it is based on intercorrelations, it has involved giving all three inventories to a single group, consisting of 200 college students.

The Inventories

These three inventories provide measures of the following factors:

¹ The writer wishes to express her gratitude to Lt. Col J. P. Guilford for his helpful suggestions concerning this research.

² J. P. Guilford and R. B. Guilford, "Personality Factors S, E, and M, and Their Measurement," *Journal of Psychology*, II (1936), 107-127, "Personality Factors D, R, T, and A," *Journal of Abnormal and Social Psychology*, XXXIV (1939), 21-36, "Personality Factors N and GD," *Journal of Abnormal and Social Psychology*, XXXIV (1939), 239-248.

C. I. Mosier, "A Factor Analysis of Certain Neurotic Tendencies," *Psychometrika*, II (1937), 263-287.

- S—Social Introversion-Extraversion (sociability, tendency to seek social contacts and to enjoy the company of others as against shyness, tendency to withdraw from social situations and to be seclusive)
- T—Thinking Introversion-Extraversion (lack of introspectiveness and an extravertive orientation of the thinking process in contrast to an inclination to meditative thinking, philosophizing, analyzing oneself and others, and an introspective disposition).
- D—Depression (freedom from depression and possession of a cheerful, optimistic disposition versus a chronically depressed mood and possession of feelings of unworthiness and guilt).
- C—Cycloid disposition (stability of emotional reactions and moods and freedom from cycloid tendencies in contrast to strong emotional reactions, fluctuations in moods, and a disposition toward flightiness and instability).
- R—Rhathymia (a happy-go-lucky or carefree disposition, liveliness, and impulsiveness as against an inhibited disposition and an over-control of the impulses).
- G—General activity (tendency to engage in vigorous overt action versus a tendency to inertness and a disinclination for overt activity).
- A—Ascendancy-Submission (social leadership vs. social passiveness).
- M—Masculinity-femininity (masculinity of emotional and temperamental make-up versus femininity of make-up).
- I—Inferiority feelings (self-confidence and lack of inferiority feelings as against lack of confidence, under-evaluation of one's self, and feelings of inadequacy and inferiority).
- N—Nervousness (tendency to be calm, unruffled, and relaxed in contrast to jumpiness, jitteriness, and a tendency to be easily distracted, irritated, and annoyed).
- O—Objectivity (tendency to view one's self and surroundings objectively and dispassionately versus a tendency to take everything personally and subjectively and to be hypersensitive).
- Co—Cooperativeness (willingness to accept things and people

as they are and a generally tolerant attitude as against overcriticalness of people and things and an intolerant attitude).

Ag—Agreeableness (lack of quarrelsomeness and a lack of domineering qualities in contrast to a belligerent, domineering attitude and an overreadiness to fight over trifles).

In the construction of the inventories, the following general procedure was used.³ Items were formulated which appeared to be diagnostic of each of the thirteen aspects of personality as defined by the previous work done. These were stated in question form, to be answered by "Yes," "?," or "No." Preliminary scoring keys were set up on the basis of the best statistical evidence at hand. The questions were administered to groups of subjects (e.g., 500 employed individuals in the case of the *Personnel Inventory I*). After the papers were scored with the preliminary keys the test of internal consistency was applied to every item. Those items which were not sufficiently diagnostic were discarded. For the remaining items scoring weights were assigned in accordance with a method devised by Guilford.⁴

Because of the possibility of faking answers to items the value of personality inventories has been questioned. Probably no satisfactory answer can be given without consideration of the purpose for which the inventories are used. Administering inventories to a group of prospective employees, who know that their chances of work depend on their responses, may be expected to yield different results from those obtained in a situation where individuals are motivated by the desire to gain additional information about their personalities.

Several investigations have been made in which students have been asked to take inventories twice.⁵ In one situation

³ The description of the construction of the inventories has been adapted from the discussions found in the manuals for the three inventories

⁴ J. P. Guilford, "A Simple Scoring Weight for Test Items and Its Reliability," *Psychometrika*, IV (1941), 367-374.

⁵ For example, R. G. Bernreuter, "Validity of the Personality Inventory," *Personnel Journal*, XI (1933), 383-386, C. Dowling, "Ability of College Students to Influence Scores on the Guilford-Martin Personnel Inventory," unpublished research study, The University of Southern California, 1944; J. A. M. Kimber, "The Insight of College Students into the Items of a Personality Test," unpublished doctor's dissertation, The University of Southern California, 1945, F. L. Ruch, "A Technique for Detecting Attempts to Fake Performance on the Self-inventory Type of Personality Test," *Studies in Personality*, New York: McGraw-Hill Book Company, 1947.

they have been requested to respond according to the way they think they are; in the other, according to the way they would like to be, the way they think a well-adjusted individual would respond, or the way they think a good employee would respond. Such studies have revealed consistently a difference in scores between the two situations. The results show that responses can be influenced in a given direction but they also give an indication that students do not answer the items, under the ordinary procedure of administration, so as to present the best possible picture of themselves. Inasmuch as the present study was conducted in a manner similar to the "normal" condition in the above investigations, it is probably not unreasonable to assume that a similar attitude toward the inventories was present.

Such findings, of importance in relation to the matter of faking answers to items, are of course not decisive evidence of the validity of the inventories. More direct information has been obtained. In one study,⁶ inventory scores for factors S, T, D, C, and R were correlated with self-ratings and with ratings by close associates. The reliabilities of the ratings for T and C were not sufficiently acceptable as criteria against which to validate scores for those factors. For S, D, and R, the correlations were high enough to indicate that the inventory scores were quite valid.

In another study, the validity of factor M was checked by comparing the distributions of the inventory scores of 50 males and 50 females not used in the original standardization group.⁷ Forty-six of the males were above the median of the distribution of the scores of the two sexes combined and forty-six of the fifty females were below the median. The validity coefficient (ϕ) for the factor was .84. It was considered highly satisfactory in view of the fallibility of biological sex as a criterion of masculinity-femininity as a temperamental trait.

With the *Personnel Inventory I*, a study was made in which workers were classified into a "satisfactory" group and an "un-

⁶ J. P. Guilford and Howard Martin. "Age Differences and Sex Differences in Some Introverted and Emotional Traits," *Journal of General Psychology*, XXXI (1944), 219-229

⁷ Description of this study is taken from the manual of directions for the inventory.

satisfactory" group on the basis of test results.⁸ The inventory was taken under conditions in which the subjects were informed that their employment status would not depend on the results. Of 22 workers judged unsatisfactory by management, 68% were detected by the test. Of 26 workers judged satisfactory by management, 73% were correctly placed by the test. Other studies have yielded results in line with this one.⁹ The authors of the inventory have pointed out that, in these preliminary studies, selection of unsatisfactory individuals was made in terms of arbitrary criteria and that more detailed study of the jobs in question might have led to the use of different cut-off points and greater success. In the manual of directions they urge that, for usage of this sort, critical scores be based on experience in the specific situation.

Reliability coefficients for the inventories have been given in the manuals. They were computed by dividing the scored items for each factor into two random halves, computing Pearson coefficients of correlation, and then estimating reliability coefficients by means of the Spearman-Brown formula. The reported reliability coefficients are as follows: S = .90, T = .84, D = .94, C = .88, R = .90, G = .89, A = .88, M = .85, I = .91, N = .89, O = .83, Ag = .80, Co = .91.

Procedure

The three personality inventories were administered, according to the directions in the manuals, to four elementary psychology classes at The University of Southern California. Before the inventories were given out an appeal for cooperation in securing accurate responses was made. The students were informed that they would be given their results individually and that the scores they made would have no influence on their grades in the course.

Two hundred and thirteen subjects completed all three inventories. They were divided, according to sex and nearest age, as follows:

⁸ R. M. Dorcus. "A Brief Study of the Humm-Wadsworth Temperament Scale and the Guilford-Martin Personnel Inventory in an Industrial Situation," *Journal of Applied Psychology*, XXVIII (1944), 302-307.

⁹ H. G. Martin. "Locating the Troublemaker with the Guilford-Martin Personnel Inventory," *Journal of Applied Psychology*, XXVIII (1944), 461-467.

Age:	15	20	25	30	35	40	45	50
Number of Men:	3	101	20	1	1	0	0	0
Number of Women:	4	71	3	4	2	1	0	2

Of these cases those whose nearest age was 15 and those whose nearest age was 35 or above were dropped. This selection left 200 cases: 122 men and 78 women. It was made because those at the extremes in age might give atypical results for college students and because the loss of such a small number would make no appreciable difference as far as statistical significance was concerned.

Raw scores for each of the factors were determined for each subject. These were converted into scaled scores by means of the conversion tables in the manuals. The scaled scores (C scores) were originally set up on the groups used in standardization to normalize the distributions for the various factors.¹⁰

Intercorrelations of each factor with every other factor were then computed, using the Pearson product-moment method. These intercorrelations are given in Table 1. Sixty-five of them were significant (at the 5% level), being .140 or greater. Sixty-two of them, .182 or greater, were very significant (at the 1% level).

The Thurstone method of factor analysis was used. Centroid factors were extracted according to the procedure given

TABLE 1
Intercorrelations of Factor Scores

	S	T	D	C	R	G	A	M	I	N	O	Ag	Co
S423	.638	.439	.655	.379	.733	.101	.591	.384	.465	.140	.222
T645	.588	.300	-.070	.197	.212	.335	.391	.405	.169	.237
D901	.228	-.040	.481	.315	.740	.710	.746	.337	.442
C	-.021	-.188	.308	.330	.675	.701	.722	.351	.416
R559	.525	.039	.270	.079	.207	-.084	-.019
G438	-.067	.088	-.231	-.059	-.314	-.169
A256	.570	.325	.460	.001	.200
M326	.348	.365	.006	.210
I674	.746	.350	.448
N720	.470	.529
O495	.616
Ag631
Co

¹⁰ J. P. Guilford *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill Book Company, 1942. Pp. 104-106.

by Guilford.¹¹ In estimating communality the highest correlation in each column was used. It was decided to continue extraction as long as the range of factor loadings was at least $-.20$ to $+.20$. This criterion called for the extraction of six factors.¹² In the following discussion these are consistently called "super-factors" to emphasize their distinction from the thirteen original inventory factors.

With these the communalities for the thirteen factors were found by computing the sum of the squares of the super-factor loadings for each. Comparison of these with the communalities estimated at the beginning of the analysis revealed one difference of $.145$, which was considered too large to be tolerated. Accordingly a second set of extractions was made using the communalities obtained from the super-factor loadings of the first extractions. This time the largest discrepancy between the estimated and obtained communalities was $.038$, which was considered well within the limits of toleration.

TABLE 2
Centroid Super-factor Loadings and Communalities from Second Extraction

Factor	Super-factor loading						Obtained communality	Estimated communality	Discrepancy
	I	II	III	IV	V	VI			
S	.761	-.477	-.166	-.092	.226	.047	.896	.876	.020
T	.560	.109	-.521	-.123	-.198	.060	.655	.617	.038
D	.896	.197	-.292	.132	.076	-.057	.953	.969	.016
C	.780	.423	-.280	.260	.089	-.125	.957	.968	.011
R	.433	-.665	-.077	-.182	-.187	.084	.711	.698	.013
G	.119	-.740	.054	.076	-.116	-.242	.643	.620	.023
A	.668	-.498	.190	.195	.081	.114	.788	.812	.024
M	.351	.153	.106	.279	-.270	.227	.360	.342	.018
I	.828	.050	.155	.180	.122	-.023	.760	.759	.001
N	.736	.379	.076	.039	.087	.166	.728	.743	.015
O	.846	.253	.188	.014	-.062	-.057	.822	.827	.005
Ag	.404	.445	.217	-.472	.188	-.116	.680	.657	.023
Co	.558	.367	.349	-.316	-.024	-.068	.673	.670	.003

¹¹ J. P. Guilford, *Psychometric Methods*. New York: McGraw-Hill Book Company, 1936. Pp. 478-488.

¹² Comparison of the standard deviations of the residuals with the standard error of the average correlation indicated that not more than three factors should be extracted. However, it is considered only a rough test. Coombs' criterion gave inconsistent results. Tucker's criterion (revised) indicated that at least seven factors should be extracted. Because of the inconsistency of these results it was decided to continue extraction as long as the range of loadings was $-.20$ to $+.20$. Beyond that point (with a maximum contribution to communality of less than $.04$) it did not seem advisable to go.

The centroid loadings from this analysis, used in the rotations which followed, are given in Table 2 together with the obtained communalities, the estimated communalities, and the discrepancies.

Rotation of the axes was made graphically, according to the procedure given by Guilford.¹³ The aim was to minimize the size and number of negative entries and to maximize the number of vanishing entries.¹⁴ Rotation was continued until no further improvement according to these criteria could be obtained. The super-factor loadings and communalities from the final rotation are given in Table 3.

TABLE 3
Super-factor Loadings and Communalities after Rotation

Factor	Super-factor loading						Communality
	I	II	III	IV	V	VI	
S	.704	.085	.422	.060	.245	.390	.8966
T	.084	.181	.625	-.053	.455	-.056	.6526
D	.233	.438	.813	.089	.093	.162	.9498
C	.017	.480	.841	.059	-.080	.090	.9560
R	.711	.045	-.027	-.061	.441	.072	.7116
G	.734	-.091	-.170	-.234	.000	-.088	.6385
A	.704	.383	.090	.001	.047	.370	.7895
M	-.017	.584	.066	-.050	.096	.038	.3589
I	.377	.537	.445	.240	-.092	.255	.7596
N	.003	.542	.488	.351	.065	.258	.7259
O	.248	.601	.471	.415	.018	.051	.8196
Ag	-.082	.060	.330	.748	.010	.012	.6789
Co	.058	.357	.250	.693	.040	-.042	.6769

Two negative loadings remained after the final rotation. Trait G had a loading of $-.170$ on Super-factor III and a loading of $-.234$ on Super-factor IV. In view of the fact that this factor had four significant negative correlations with other factors, failure to achieve a positive manifold through rotation is not unreasonable. Negative loadings of G on these super-factors, moreover, fit logically into the interpretation given to them.¹⁵

¹³ J. P. Guilford, *op. cit.*, 489-491 and 502-507.

¹⁴ Loadings of $+ .3$ and above or of $-.3$ and below were considered significant in naming super-factors; those from $+ .11$ to $+ .29$ and from $- .11$ to $- .29$ were considered as different from zero but too small to be important in identification, and those between $+ .10$ and $- .10$ were regarded as vanishing.

¹⁵ See data describing Super-factors III and IV.

Interpretation of Results

Listed below are the loadings of the thirteen factors on Super-factor I, in order from highest to lowest:

<i>Factor</i>	<i>Loading</i>	<i>Data describing Super-factor I</i>
G	.734	Has tendency to engage in vigorous overt action
R	.711	Happy-go-lucky, lively, impulsive, uninhibited
S	.704	Sociable, has tendency to seek social contacts and to enjoy company of others
A	.704	Tends to be social leader
I	.377	Self-confident
O	.248	Objective
D	.233	Cheerful, optimistic
T	.084	May or may not be introspective
Co	.058	Either tolerant or intolerant
C	.017	May have stable or unstable emotional reactions
N	.003	Either relaxed or nervous
M	-.017	Either masculine or feminine in emotional make-up
Ag	-.082	May or may not be quarrelsome and domineering

This super-factor has been identified tentatively as a drive-restraint variable. Those factors with sizable loadings on it appear to have in common an active approach to experience. The person with high scores on them tends to engage in *vigorous overt action*, to give relatively *uninhibited expression* to impulses, to *seek* social contacts, and to be a *social leader*. This super-factor gives the contrast between the individual who pushes out into activity as against the person who has to be forced into it.

The other positive loadings, though not high enough for use in naming the super-factor, are in agreement with the identification made. One might expect that drive for response would tend to be accompanied by feelings of confidence in and optimism about reactions made, and that pressure for response might prevent an individual from becoming prey to hypersensitive reactions.

Moreover, the vanishing loadings seem in accord with the identification. It appears logical to think of drive for response as being independent of degree of tolerance, emotional stability, nervousness, masculinity of make-up, and domineering tendency. The only loading that is difficult to fit into this picture is the vanishing one on T. However, if one thinks of T in terms of the differentiation it makes between extravertive and introvertive orientation of the thinking process rather than in terms

merely of tendency toward meditation, the vanishing loading seems more reasonable.

Below, in order of size, are the loadings on Super-factor II.

<i>Factor</i>	<i>Loading</i>	<i>Data describing Super-factor II</i>
O	.601	Objective
M	.584	Has masculine attitudes
N	.542	Calm, unruffled, relaxed
I	.537	Self-confident
C	.480	Has stable emotional reactions
D	.438	Cheerful, optimistic
A	.383	Is social leader
Co	.357	Tolerant
T	.181	Lacks introspectiveness
S	.085	May or may not be sociable
Ag	.060	May or may not be quarrelsome
R	.045	May or may not be happy-go-lucky
G	-.091	May or may not have tendency to engage in vigorous overt action

This super-factor has been tentatively named a realism variable. The inventory factors with high loadings on it present a good picture of the impersonal and dispassionate realist. He views things objectively. He does not go to pieces at seeing a fish on a hook. He is calm, unruffled, and self-confident (for he is objective enough to know that his bad points aren't his whole personality). Also, because of his objective and impersonal approach, he tends to have stable emotional reactions and not to become unduly depressed by passing disappointments. One might expect that such an individual might have some tendency toward tolerance and leadership, though the relatively low loadings of these factors are not unreasonable in light of the identification. The vanishing loadings present a logical addition to the description of this super-factor. It seems reasonable to think of this characteristic of realism as being independent of degree of sociability, impulsiveness and carefreeness, tendency to engage in vigorous overt action, and tendency toward quarrelsomeness.

This super-factor presents a fairly good picture of reported sex differences in personality except for the low loading of social leadership, which is supposed to be more characteristic of men than of women. However, it seems preferable to name the variable in terms of the attitudes and reactions it involves rather than to call it simply "masculinity-femininity."

The loadings on Super-factor III, in order, are as follows:

<i>Factor</i>	<i>Loading</i>	<i>Data describing Super-factor III</i>
C	.841	Has stable emotional reactions
D	.813	Is cheerful, optimistic
T	.625	Lacks introspectiveness
N	.488	Is calm, unruffled
O	.471	Objective
I	.445	Self-confident
S	.422	Sociable
Ag	.330	Lacks quarrelsomeness
Co	.250	Tolerant
A	.090	May or may not be a social leader
M	.066	May or may not have masculine attitudes
R	-.027	May or may not be carefree
G	-.170	Less active than the average person

This super-factor has been defined tentatively as an emotionality variable. At the low extreme on it would be the individual characterized by hampering emotional excess. At the other extreme (as indicated by the high loadings) would be found the individual who is dependably cheerful and optimistic, free from constant analysis of himself and others, with some tendency to be (1) free of nervous habits, (2) lacking in hypersensitivity, (3) self-confident, sociable, and tolerant, and (4) lacking in domineering qualities. Such an individual might or might not be a leader in social situations, masculine in his attitudes, and uninhibited. It is logical to think that he might have some tendency to be a "slow mover," since a person with great drive for activity would be likely to get into more upsetting situations. However, the negative loading on G is not large enough to merit much consideration in the naming of this super-factor.

For Super-factor IV, the following loadings were found:

<i>Factor</i>	<i>Loading</i>	<i>Data describing Super-factor IV</i>
Ag	.748	Lack of quarrelsomeness and domineering qualities
Co	.693	Tolerant
O	.415	Objective
N	.351	Calm, unruffled, relaxed
I	.240	Self-confident
D	.089	May or may not be optimistic
S	.060	May or may not be sociable
C	.059	May or may not have stable emotional reactions
A	.001	May or may not be a social leader
M	-.050	May or may not have masculine attitudes
T	-.053	May or may not be introspective
R	-.061	May or may not be carefree
G	-.234	Less active than the average person

This super-factor has been identified tentatively as a social adaptability variable. The factors with high loading on it seem to present a picture of the individual whose actions are influenced by the desire for smooth relations with others. He does not domineer over others or quarrel with them; he is tolerant of others' beliefs; and he is objective in his interpretations (such objectivity being necessary for smooth adjustment to other people). Perhaps because such a person adapts himself to others easily he tends to be calm and relaxed and to be self-confident.

The vanishing loadings fit readily into this picture. The person who is concerned with adapting himself to the responses of others may or may not be (1) cheerful, (2) desirous of going out of his way to seek social contacts, (3) high in leadership qualities (he might be either a good leader or a good follower), (4) masculine in attitudes, (5) introspective, (6) impulsive, or (7) stable in mood. Further, one might expect that such a person would tend to have rather low pressure for overt activity, since it would make for fewer chances of disagreement with others.

Super-factor V had the following loadings:

<i>Factor</i>	<i>Loading</i>	<i>Data describing Super-factor V</i>
T	.465	Lacks introspectiveness
R	.441	Carefree, impulsive
S	.245	Sociable
M	.096	May or may not have masculine attitudes
D	.093	May or may not be cheerful
N	.065	May or may not be calm and relaxed
A	.047	May or may not be social leader
Co	.040	May or may not be tolerant
O	.018	May or may not be objective
Ag	.010	May or may not be domineering
G	.000	May or may not be vigorous in action
C	-.080	May or may not have stable emotional reactions
I	-.092	May or may not be self-confident

Below are the loadings for Super-factor VI:

<i>Factor</i>	<i>Loading</i>	<i>Data describing Super-factor VI</i>
S	.390	Sociable
A	.370	Social leader
N	.258	Calm, relaxed
I	.255	Self-confident
D	.162	Cheerful, optimistic

C	.090	May or may not have stable emotional reactions
R	.072	May or may not be carefree, lively
O	.051	May or may not be objective
M	.038	May or may not have masculine attitudes
Ag	.012	May or may not be domineering
Co	-.042	May or may not be tolerant
T	-.056	May or may not be introspective
G	-.088	May or may not have tendency toward vigorous overt action

These two super-factors are too weak to be of any importance. Both are merely doublets, accounting for special correlations (in addition to the influence of the other super-factors) between S and A and between T and R.

One finding of particular interest in this study is the fact that no very general super-factor was located which might be called "tendency to give the desirable response" or "insight into the desirability of the response." Opinion as to just what score on these thirteen factors a very well adjusted person should possess would vary somewhat from individual to individual. However, probably most persons would agree with the authors of the inventories that the following scores are desirable: high scores on S, D, C, A, I, N, O, Co, and Ag; middle scores on T, R, and G; and a score on M depending on sex. If individuals were answering the items in terms of their insight into the desirability of the items, one would expect to find a super-factor in which S, D, C, A, I, N, O, Co, and Ag had sizable loadings. Nothing approaching this was found. Apparently understanding of the desirability of certain responses did not have a marked influence on results. This finding is in line with the material cited early in the report concerning normal and special methods of administering inventories.

The results of this study present interesting suggestions concerning the structure of personality. On the basis of the findings one may conceive of personality as consisting of hierarchies of habit systems of different degrees of independence and generality. The smallest units are the habit systems tapped by individual items of the inventories. Many of them are inter-correlated. They fall into clusters because they have in common some more general characteristic. These characteristics are not only less specific but are on the average more independent of each other. (Such are the thirteen factors measured by

these three personality inventories.) They, in turn, are inter-correlated to a certain extent. They fall into certain clusters because of even more general factors they have in common. These super-factors are more separate from each other on the average than the less general habit systems.

More particularly, this study has indicated the following four general habit systems: drive, emotionality, realism, and social adaptability. In an orthogonal structure such as this, a person may stand at any position on the scale for any of these four factors. He might, for example, be high in social adaptability, low in realism, low in emotionality, and average in drive. A person with a moderately high score on social adaptability would tend to score high on both tolerance and agreeableness (the more specific habit systems which have this characteristic in common), because the two are positively correlated. However, these correlations are low enough so that, in individual cases, there might be considerable disparity between standings on the two. Therefore separate scores for each are indicated. These, of course, are the factor scores from the inventory.

For a more concise and more generalized picture of an individual's personality than that provided by the thirteen factor scores one would want measures of the four super-factors. Equations for predicting such scores from the thirteen C scores have been set up using the Doolittle method. In this process an arbitrary mean (50) and an arbitrary standard deviation (10) for the super-factor scores have been assumed. Moreover, only those traits with super-factor loadings of .5 or above have been used. These prediction equations are as follows:¹⁶

$$\begin{aligned} I &= 28.370 + 1.78G + .682R + .851S + .891A. \\ II &= 36.101 + .804O + 1.202M + .349N + .299I. \\ III &= 30.894 + 2.273C + .745D + .569T. \\ IV &= 33.967 + 1.805Ag + 1.339Co. \end{aligned}$$

In addition to the above, one further result should be mentioned. The original studies made by Guilford indicated that factors D and C were sufficiently independent to warrant separate measurement. Items were then constructed which appeared to be measuring each. Obviously these items were not

¹⁶ The following multiple correlation coefficients were obtained. $R_{I-GHSA} = .888$, $R_{II-OMNI} = .729$; $R_{III-ODT} = .860$, $R_{IV-AGCo} = .800$

pure measures, for the correlation obtained in this study for the scores on the two factors was .90. This indicates that additional work on these two sets of items is necessary to bring the correlation between scores on the inventory closer to the correlation of the factors themselves, as found in the original research. There were a number of other correlations in the seventies. These were for factors in separate inventories for which no correlations are available. It might be possible to lower these somewhat by the removal of impure items. However, in view of the general interpretation of the results of this study, one would not expect to eliminate all correlation even if perfectly pure items for each factor were used. And, as they stand now, the correlations are not high enough to enable accurate prediction of one factor from the other.

The results as given are, of course, limited by the selection of subjects and the procedure used in the study. Generalization of these findings for college students to all individuals is not warranted. Further research, set up in similar form, should be done with non-college groups as subjects. In addition, the findings would be expected to apply only in cases where the inventories were given under the conditions of administration used in this investigation. One would predict different results, for example, if subjects were asked to take the inventories so as to indicate how a happy, well-adjusted person would respond. Factor analysis of scores obtained with such a procedure would make an interesting study.

Summary

The purpose of this study was to make a factor analysis of the thirteen variables of personality measured by Guilford's *Inventory of Factors STDCR*, the Guilford-Martin *Inventory of Factors GAMIN*, and the Guilford-Martin *Personnel Inventory I*.

The three inventories were administered to two hundred college students under standard conditions. The results obtained in the study are, of course, limited by these selective factors.

For each of the subjects scaled scores were obtained on the

following factors: sociability, extravertive orientation of the thinking process, freedom from depression, stability of emotional reactions, carefreeness, general drive, social ascendance, masculinity, freedom from inferiority feelings, freedom from nervousness, objectivity, lack of quarrelsomeness, and tolerance. Intercorrelations between the scores were then computed and a factor analysis of the results was made, using the Thurstone method. Six super-factors were obtained. The first four were identified tentatively as:

- I. Drive-restraint (high loadings on general drive, carefreeness, sociability, and social ascendance).
- II. Realism (high loadings on objectivity, masculinity, freedom from nervousness, and freedom from inferiority feelings).
- III. Emotionality (high loadings on stability of emotional reactions, freedom from depression, and extravertive orientation of the thinking process).
- IV. Social adaptability (high loadings on lack of quarrelsomeness and tolerance).

The remaining two super-factors were doublets, accounting for special relations between two pairs of factors. No super-factor which seemed to involve insight into the desirability of the responses or tendency to give "good" responses was found.

For the subjects used the results seem to picture the structure of personality as consisting of four general areas, relatively independent of each other, within which lie less general habit systems (less independent, but sufficiently so to make separate scores advisable for diagnostic work). Equations for predicting scores on the four super-factors from the thirteen factor scores were set up using the Doolittle method.

The high correlation between scores on C (stability of emotional reactions) and D (freedom from depression) indicated the advisability of revision of these two sets of items to bring them closer to the correlation originally found between the factors themselves.

THE OFFICE OF RADIO RESEARCH:¹ A DIVISION OF THE BUREAU OF APPLIED SOCIAL RESEARCH, COLUMBIA UNIVERSITY

MARJORIE FISKE AND PAUL F. LAZARSFELD

Office of Radio Research

THE rapidly increasing development of the radio industry in the past two decades has opened up a new and increasingly important area of communications research. Radio is accessible to more people than any other kind of communication. Its effects are therefore of increasing importance to the sociologist, the educator and the politician. And since radio in America is a privately owned and operated industry, its impact is also a matter of importance to networks, advertisers, advertising agencies and others concerned with its commercial effectiveness.

The Office of Radio Research has been functioning for seven years, during which it has been concerned with a wide variety of radio research problems. In some instances these problems could be solved by a relatively simple adaptation of techniques used in other fields of research. In others it was necessary to develop new techniques to fit the special characteristics of this relatively new medium. It will not be possible within the scope of this chapter to enumerate all such adaptations and innovations. We shall, rather, touch upon a few which illustrate the interrelationship of radio and other fields of communications research.²

The Objectives of Radio Research

It is manifestly impossible to study either the content or the effect of every radio broadcast that goes on the air. The

¹ This article is a chapter of a book, *How to Conduct Consumer and Opinion Research*, edited by A. B. Blankenship, which is to be published by Harper and Brothers early in 1946.

² The authors are indebted to Dr. Bernard Meyers for his assistance in organizing this material.

larger body of knowledge, the over-all picture, has to be built up over a period of time, segment upon segment, each segment representing a study of a particular program or a particular group of listeners. Furthermore, studies of particular programs or particular groups of listeners may be done from the different standpoints of the educator, the politician, the sociologist, the psychologist and the businessman, each seeking answers to different questions. Rarely is one program or one group of listeners studied from all of these viewpoints, but each study contributes to the sum total of knowledge of radio's rôle in our culture. Each contributes new techniques which can be used by the other. Altogether, they are gradually being interwoven to form an increasingly important part in the general pattern of communications research.

Some Examples of the Several Approaches to Radio Research—The Office of Radio Research has had occasion to do research from all of these several standpoints. In studies reported in *Radio and the Printed Page* (11) radio was viewed from the standpoint of the educator; reading and listening habits were compared and new insights gained as to the relative rôles of radio and print as informational media. It has studied a particular program, the Orson Welles "Invasion from Mars" broadcast, for example, from the standpoint of the psychologist to gain insight into how different kinds of people react to a given stimulus situation (3). This broadcast, it will be recalled, resulted in near panic in certain parts of the country. To determine how it came about that a radio drama could spread genuine terror through a substantial part of the population, detailed interviews were made with many different kinds of listeners who reacted in a variety of ways. The results not only shed light on the power and potentialities of radio as a medium of communication but enabled the psychologist to gain insight into the general psychology of panic.

Again, certain programs and certain kinds of listeners have been studied from a broad sociological standpoint as in the not yet published study⁸ of the Kate Smith all-day bondselling appeal, wherein a particular broadcast of a popular entertainer

⁸ "Swayed By Smith" A chapter in *The Social Psychology of Mass Persuasion*. Robert K. Merton, with the assistance of Marjorie Fiske and Alberta Curtis. To be published by Harpers early in 1946

was subjected to scrutiny. In this study the content of the program was analyzed to determine the variety of its appeals and the listeners were interviewed at length to determine the relative impact of these appeals, and how they came to decide to buy a bond from Kate that day. In the course of the analysis it became apparent that a particular radio entertainer may epitomize the social force of radio, reflecting certain trends and concepts in our culture, and at the same time reinforcing them. Kate, for example, lays great stress on the sacrifices and sacredness of motherhood. Indeed for many of her listeners she has come to typify motherhood: "Only a mother could plead the way she does," even though most of them know she is not a mother. In this extolling of motherhood she not only reflects one of the basic concepts of our time and our culture, but at the same time she reinforces and strengthens it.

The program sponsor and the advertising agency study radio from yet another angle. They are concerned with the extent to which their "messages" get across and with the extent of acceptance or rejection of their particular programs. Here the research focuses on the immediate response-reactions of the listeners. An effort is made to determine the extent to which a program is liked or disliked and why, and to find out its effect on the subsequent attitudes or actions of those who heard it. Studies of this kind involve either program research or the testing of commercials, techniques which we shall consider shortly.

Viewing radio as a whole it is apparent that its content is the product of contemporary culture and customs. Its content reflects this culture because the people responsible for it are in turn the products of it. *Analyses of the content* of radio broadcasts, therefore, provide the sociologist with a better understanding of society. On the other hand, this content has an impact on millions of radio listeners, and by studying the effects of radio on listeners' habits and attitudes the sociologist also gains insight into the way in which it is changing, modifying or reinforcing the cultural pattern.

The Techniques of Radio Research

Surveys of General Listening Habits.—The most general kind of listener survey involves a simple count of how many

people listen to a given program. Such counts, known as program ratings, are based on careful samplings of the population and are made systematically by a number of research organizations. Program ratings and fluctuations are thus made available regularly to private clients.⁴ The Office of Radio Research, however, while it does make use of such ratings in some of its more quantitative studies, e.g., *"The Social Stratification of the Radio Audience"* by Hugh Beville (2), confines itself largely to studies of a more detailed psychological nature.

Surveys of general listening habits may be made for several reasons. One may wish to compare the influence of the various media of communication on a given area of behavior. How, for example, does the influence of radio compare with the influence of newspapers and magazines on voting behavior? (13). Or one may want to measure changes in listening habits resulting from program changes, or to gain insight into the rôle of radio among certain groups of people. In this case one must also survey general listening habits over a period of time. Whatever the purpose of such general surveys, the procedure is the same: something akin to a "listening diary" must be procured from a representative sample of the population one wants to study.

Several studies of this kind, both commercial and non-commercial, have been made at the Office of Radio Research. A good example of this is the question of listening in the daytime. If the whole listening pattern of women were taken into consideration they would fall into three equally large groups. Day-time Listeners include those who listen to serials and those who listen to other programs. Those who are at home in the morning and could listen but do not comprise the Non-Listeners. Thus radio actually does not reach a third of the available morning audience and, at the same time, has to cater to two rather different kinds of audiences. How to reconcile the interests of these two divergent sectors of the audience is a problem which leads to a large number of interesting and still partly unsolved research problems.

Another survey of general listening habits was concerned with the question of who listens to small local stations. This

⁴ See Chapters XI, XII, XIII, and XIV.

involved interviewing a cross section of the radio audience in a given locality about their radio listenership over a period of time, and comparing different age, sex and socio-economic groups in respect to station preferences (16). It developed that people on the lower socio-economic strata tend to listen to such stations more than do those who are better educated and better situated financially.

A study of a somewhat different nature recently completed by the Office also falls into this category. Here the problem was to determine (a) the degree of satisfaction with current radio offerings, (b) attitudes toward commercial advertising on the radio, and (c) general receptivity toward a proposed new plan by which the listener would subscribe to a service which would provide him with three types of radio programs without any advertising.

Still another kind of general listenership survey is designed to determine the rôle of radio in the lives of particular groups—children, for example, or housewives, or certain socio-economic groups. Such a study usually involves careful and detailed case histories of representatives of the group under study. This kind of investigation is well exemplified in two Office of Radio Research studies, "Listeners Appraise a College Station" (4), and "Radio Comes to the Farmer" (19). In the latter, it was possible, by use of the detailed interview method, to determine the extent to which the acquisition of a radio changed the habit and thought patterns of a group of Iowa farm households.

A similar type of survey is indicated when a sponsor or a group of sponsors wants to measure the impact of his radio appeals or to compare it with appeals in other media (12). It was found possible in one study, for example, to gauge both the actual results of radio advertising and to get some idea of its potentialities (6). The investigators went into a representative, moderate-sized community and interviewed several hundred housewives at great length about their listening habits, their awareness of retail merchandising over the air, their varying degrees of receptiveness toward the various kinds of retail advertising and the extent to which such attitudes influence their buying habits. Among other things this study indicated

that there are certain kinds of programs which are better adapted to the selling of retail merchandise than others, and that certain kinds of merchandise lend themselves better than others to advertising over the air.

In the course of such studies it has become clear that non-listeners are also an important factor in radio research, both from the standpoint of particular programs (17) and from the standpoint of non-listening in general. If we know why certain people do not listen to a given type of program and how many people do not listen for these reasons, we can plan program changes which may not only improve content level but at the same time increase the total amount of listenership. Similarly, if an extensive survey were made of people who seldom or rarely listen to the radio at all, we could round out our picture of radio as a cultural expression and a cultural tool.

The Nature of Program Research.—Such broad audience surveys as those outlined above cannot possibly encompass the more specific problems of listener likes and dislikes, listener gratifications or the extent to which listener attitudes are changed or modified by radio listening. Therefore, the investigator finds it increasingly necessary to study particular programs or series of particular programs. In doing so, however, he comes face to face with research problems which are not especially peculiar to radio but which have their counterparts elsewhere in the communications field. How do you measure listener reaction to a program? What specifically does a listener mean when he says he liked or disliked a certain program? How can one measure the "effectiveness" of a given informational program? How determine the cumulative effect of a series of programs?

There are three different ways of learning what a program means to people: by subjecting the program to a content analysis, by making a differential analysis of the personal characteristics of the groups that listen to the program, or by asking people directly what the program means to them. Wherever possible all three methods should be used simultaneously.

Content analysis of radio material involves essentially the same techniques as are used in the analysis of printed materials,

and are usually based on scripts or transcripts of the broadcasts (unless one is studying all programs for the occurrence of a certain type of a content in which case one has to resort to "monitoring"). From analysis the investigator is able to list most of the affective factors of a broadcast. Thus the content analyst, after listening to a few instalments of a daytime serial script, may learn that it stresses an individualistic, competitive type of social relationship, that the surgeon hero wants to be a great man and stand high in a prestige hierarchy, that his interest is in himself and not in humanity. Or he may discover that the negro character depicted in the series is a servant whose chief characteristic is doglike devotion to his master with little or no portrayal of any individual thoughts, feelings or individuality of his own. In another study a content analysis of a Kate Smith script reveals that she sometimes uses the word America or American as many as seven times in five minutes, thus building up an associative complex which contributes to her reputation as a patriot.

But, as we have already suggested, content analysis is important largely as the first step in the study of any particular program or series of programs. In subjecting the script or scripts to a preliminary content analysis, the investigator accomplishes two objectives. He is able to distribute his questions on the various components of the broadcast with some regard to their frequency and importance. Secondly, a thorough-going content analysis permits certain inferences about what the listeners may get out of the content, or at least will give the investigator some idea of what *not* to look for. Because it provides both balance and perspective, content analysis is usually the first step in program research, whether it be an investigation of one program or a series of programs.

The second way to find out what a program means to people is to determine what sex, age, and social groups listen to it. Much is known about the psychological differences among various strata of the population, and if the program is listened to by some of this group more than by others, the nature of its appeal can be more readily understood. If, for example, the audience of one of two comedians is more highly educated than

the audience of another, then it can safely be assumed that the first offers a more sophisticated kind of humor. The characteristics which are to be isolated will of course vary with the problem at hand. In a study of the audience for a child guidance program, for example, whether or not the listener has children was a pertinent factor. It was found here that quite a number of childless women were found among the regular listeners, hence the conclusion that the practical advice offered is not the only appeal in this program. Some women, regretting their lack of children, might derive a vicarious satisfaction from hearing child problems discussed, while for others the broadcasts might have general educational value.

If more general listening habit surveys included detailed information about the listener, such as reading habits, leisure time activities, community participation and so on, such material would become a useful tool for the further analysis of what certain kinds of programs mean to listeners.

One of the major problems of program research is how to get the respondent to indicate what, in the program or series of programs under study, is responsible for his reactions to it. It means little or nothing to the program planner, for instance, if he is told that 70% of the respondents studied liked a program very much, 20% liked it moderately well and 10% did not like it at all. It does not tell him what could be done to improve the attitudes of the other 30% or whether changing it to meet their taste would not at the same time antagonize the 70% who liked it in its original form. After considerable experimentation, however, the Office of Radio Research has developed a technique which seems to contribute to the solution of this basic problem. This technique involves the adaptation of the polygraph frequently used in experimental psychology, and is known as the Lazarsfeld-Stanton Program Analyzer.

The Program Analyzer is an apparatus which enables a group of respondents to record their reactions to a radio program, as they listen to it, by pressing red ("dislike") and green ("like") buttons, or by not pressing buttons, which signifies indifference to what is being heard. The push buttons are connected with a pen which moves along a roll of tape synchronized

with the radio program, thus making a permanent record of the reactions of the group (8, 15).

Such a record alone, while interesting as a picture of the high and low points of the program as far as a given group of listeners is concerned, is comparatively meaningless from the standpoint of program improvement. The important thing for the program planner or the educator to know is what there was about a particular part of the program that the listener found dull or interesting and *why* in terms of the listener's own background and experience. The Program Analyzer Technique is therefore nearly always combined with a *focused interview*⁵ in which the trained investigator, using the Program Analyzer graph for reference, is able to determine just what it was about the program that caused the reactions indicated on it, and just what these reactions mean in the experience of the listener. Every research man who has tried to determine the "why" of reactions to a particular experience will recognize the advantage of this method. It gives him a picture of reactions which occurred simultaneously with the experience, and obviates a frequent difficulty in retrospective interviewing, to wit that the respondents often fail to remember how they felt in the earlier parts of the experience. Like many techniques developed in one field of communications, this one is useful in others as well and has been used successfully by the Bureau of Applied Social Research in testing reactions to motion pictures.

The Program Analyzer, of course, can be used to study reactions to any kind of program. It has been found useful in determining the effectiveness of educational broadcasts, in analyzing the appeal of entertainment programs and measuring the impact of commercial announcements. The usual procedure is to interview 10 or 20 groups of people (10 to 15 in a

⁵ The focused interview is a term applied to the technique of determining reactions to a particular communication or experience (a motion picture, a radio program, printed material and so on), known to the investigator, as distinguished from the more diffuse type of interview which is required when studying listening habits or attitudes which may be the result of several different experiences which are usually unknown to the investigator. The focused interview is a rather complicated procedure, and the O.R.R. is now in the process of codifying the results of its experience with this technique with various media of communication. The results of this systematization have been summarized by Robert K. Merton and Patricia Kendall, in an article "The Focused Interview" to appear in the *American Journal of Sociology* in the spring of 1946.

group), carefully selected to be representative of the audience the program is designed to reach. They first listen to the program, recording their reactions with the Program Analyzer push buttons, and are then questioned by a highly trained interviewer with all remarks recorded by a stenotypist. Their comments are then analyzed in conjunction with Program Analyzer graphs, and the investigator can thus determine what the effective components of the broadcast were and can make recommendations as to which parts of the program should be taken out, changed, or eliminated for more satisfactory results.

When a radio investigator wants to probe deeper, to determine the gratifications of certain segments of the radio audience, interviews of a more elaborate kind are in order. Such studies usually involve two steps. detailed and exploratory case studies, followed by less detailed interviews with a larger sample, for statistical verification of the hypotheses developed from the qualitative data. This combination of qualitative and quantitative research has two advantages. On the one hand, the first step enables the investigator to gain rich psychological insights which permit him to cover the wide range of possible responses in the statistical survey. On the other hand, the qualitative material enables him to understand, clarify and illustrate the quantitative data more fully. The combination of the two types of research has proven so fruitful that it has become an established procedure in many of the studies undertaken by the Office. Perhaps the best way to illustrate its value is by way of a concrete example.

An Example of Program Research.—The problem was to determine the gratifications of the millions of women who listen to the serial stories broadcast throughout the day by the major networks. As a first step, 100 women from various age and socio-economic groups were interviewed intensively (9). Analysis of their reports about their listening experiences and the satisfactions they derive from it indicated that there are three major types of gratification in listening to daytime serials. Some listeners enjoyed them primarily as a kind of emotional release. Burdened with their own problems, they claimed it "made them feel better to know that other people have

troubles, too." A second and more obvious form of enjoyment of the serials comes from the vicarious experiences they supply. A third gratification was entirely unanticipated by the investigator and constitutes a good illustration of the value of this kind of intensive interviewing. Many women listened to serials because they provide standards of value and judgment and help them to solve everyday problems. They learn things from these stories which they use later in solving their own problems: "Bess Johnson shows you how to handle children. She handles all ages. Most mothers slap their children. She deprives them of something. That is better. I use what she does with my own children." Or they provide comfortable philosophy for use with one's self or others: "When Clifford's wife died in childbirth the advice Paul gave him I used for my nephew when his wife died."

In this way, by providing such "leads," the intensive interviews opened up the areas for investigation on a more quantitative basis. Later, when 2,500 listeners were interviewed (10), 41% claimed to have been helped by daytime serials, thus giving statistical validity to a gratification which might have been overlooked altogether had the intensive interviews not been made. With the larger sample it became possible to make cross-tabulations which showed what kind of women found the serials helpful in this way. Thus, for example, it developed that the less formal education a woman has the more help she derives from the serials. The quantitative material also made it possible to analyze the nature of this help, and it developed that listeners find these programs useful in several ways: getting along with people, helping people with their personal problems, learning how to handle themselves in particular situations, learning how to accept misfortune with a smile, and so on.

Analysis of listener reaction combined with content analysis (1, 7, 14) of the scripts themselves then led to certain inferences about the rôle of such programs in our culture. It was found, for example, that these so-called true life stories do not deal with basic social or economic problems. They do not show a woman how she can improve her economic status nor do they give her a better understanding of the current problems

of our time—e.g., minority groups, etc. They tend, rather, to imbue the listeners with a fatalistic philosophy of life: this is how it is, we aren't as badly off as we might be. They help the listener to accept her fate by universalizing it—e.g., “husbands *never* understand their wives.” Thirdly, they encourage the listener to live life through ready-made formulas for behavior rather than helping her to develop a critical sense which will enable her to determine what is good or bad for her in a particular situation.

The field of program research, however, has been by no means completely explored. Little, for example, is known about the maximum potential of radio from an educational and cultural standpoint. We know, to be sure, that by and large the programs that are known and promoted as “educational” reach a relatively small proportion of the radio audience, chiefly those who would make a point of acquiring the same information from another medium if it were not available to them over the air. It is known that such programs will not reach even these relatively few listeners unless organized efforts are made to build an audience (14). But what about the utilization of such already accepted programs as the daytime serials as a means of raising, rather than catering to, the cultural level of the average listener? The sponsor feels he would thereby lose some of his audience. But the fact remains that few have tried to improve them and there is as yet no proof that the sponsors are right or wrong.

Effect Studies.—Another more specialized form of radio research pertains to the effectiveness of one section or element of a program. The commercial sponsor may want to determine the effectiveness of his commercial announcement. He may want to compare the effectiveness of two or more different presentations. The program planner may want to determine the extent to which his program depends upon the popularity of any single feature in it. He may want to compare commentators or announcers to determine which one is most acceptable to the greatest number of listeners. In all these cases the research procedure, as in most stimulus-response studies, involves holding all factors constant except the one under

study. Thus, to determine the relative appeal of two commercials, matched groups of respondents (or sometimes the same respondents) will listen to two broadcasts which are alike in all respects except the commercial. If there are no extraneous factors involved all differences in reaction to the two will be the result of differences in the appeal of the two commercials.

If the sponsor does not have two or more specific commercials which he wants to compare, but wants, rather, to determine the effectiveness of a particular one, the problem is somewhat different. In the first place, he must decide whether he wants to measure the effectiveness of the commercial in terms of the number of sales of his product which it induces or is likely to induce, or whether he is concerned only with the extent to which the commercial is liked or disliked. (The relationship between liking a commercial or any other kind of persuasive appeal and being induced to act as a result of it is, incidentally, a problem which needs much further exploration. Studies done to date indicate that one may dislike a commercial intensely—"spot" announcements, for example, or singing commercials—and still be influenced by them.) If the investigator is primarily interested in sales effects rather than in what elements of the commercial make it effective, a controlled check is commonly used. A section of the population is "exposed" to the appeal and sales figures for the product in that area are checked against those in a comparable area where the population was not so exposed. An alternative to this procedure involves interviewing buyers of the product to determine how they came to buy it. Most advertisers, however, seem to operate on the theory that there is a connection between liking a commercial and buying the product which it extols. Consequently they are interested in research which will determine the degree of acceptance or rejection of the commercial announcement itself. This problem involves quite different techniques.

If the advertiser is concerned only with the interest aroused by the commercial in a given program context, the Program Analyzer technique is in order. The graph will show clearly the relative position of the commercial within the framework

of reactions to the program as a whole. From the focused interview which follows he can then learn much about what was liked or disliked about the commercial and what in terms of the listeners' own experiences caused the favorable or unfavorable reaction. This technique is also useful in determining the effectiveness of commercials placed at various stages of the program—e.g., is the commercial placed at the beginning, end or middle of the program more effective? Should it follow a peak of interest in the program to capitalize on the high degree of attention at that point, or would such an approach cause a "let-down" on the part of the audience which might boomerang with resentment that "something is being put over" on the audience?

Another technique for studying commercial announcements has been found especially useful in testing reactions to "touchy" subjects, very personal products or in testing institutional advertising. This involves an intensive "depth" interview which is made immediately after the subject has read or heard the advertisement, and is equally useful for printed advertisements. Here the interview is customarily of an associative nature. What words or ideas are taboo? What words cause unpleasant associations which might in turn result in an unfavorable attitude toward the product or sponsor? This technique, incidentally, has suggested the interesting possibility that certain matters can be discussed in print which are not acceptable over the air and that contrariwise, some approaches are more effective orally than in print.

By and large, however, it has been found that the best way to make people articulate about commercial announcements, a subject which often leaves them lethargic at best, is to have them compare two or more. Most people are not sufficiently interested in such matters to become very talkative about their reactions, and the necessity of making a choice between two or more often provides the necessary impetus to self-examination as to why they selected one or the other.

Another type of program research problem to which we have already alluded involves the program series. To investigate just one of a series of educational or entertainment or dramatic

programs would not be a valid test of the effectiveness of the series, for reactions to a given program may be partly predicated on remembrance of what went before and expectation of what is to come. Then, too, if the investigator is interested in changes of attitudes as a result of a program series he will get little from merely testing one program. The technique which has been developed to solve this problem is called the "panel," which, reduced to its barest essentials, involves the selection of a group of people who agree to listen to a series of broadcasts and then report their reactions to the various programs. They may agree to come to a given place in a group and participate in a group interview using the Program Analyzer, or they may agree to record their reactions to various programs on formal questionnaires (the latter, of course, is in order when a nationwide sampling is desired). Thus a virtually constant and identical group is made available for the examination of a series of programs, and detailed comparisons of one program with another in a series can be made. Obviously, one can also procure other information from such a group—reading habits, program preferences, movie attendance, and so on, which is helpful in evaluating variations in listener reactions

Special Characteristics of Radio.—There are at least six characteristics of radio which distinguish it from other media. Researchers in the Office have had to take these into consideration when studying its effectiveness from any particular standpoint. Each of these qualities has both positive and negative aspects from the standpoint of effective communication

Perhaps most significant of these characteristics is radio's *accessibility*. Nearly every person in the United States has access to a radio. There are few geographic or economic barriers to its use once the initial investment has been made. In a sense, then, radio is more readily available than the other mass media, for each magazine, newspaper and motion picture must be purchased separately. But in another sense radio is less accessible than these other media. Once one has bought a newspaper or magazine one can keep it. It may be read at any time and an interruption or a lack of comprehension of a pas-

sage are not serious matters, for it can always be re-read. But a radio program is as ephemeral as time itself. If the telephone or doorbell rings just when John Kieran is about to answer a baseball question or just when the mystery is about to be solved one cannot set back the needle to pick up what has been lost. Motion pictures are also ephemeral in this sense, but the circumstances under which they are seen tend to offset this factor: one is not likely to be interrupted in a theater, and if a moviegoer so desires, he can always sit through a second showing.

Another special characteristic of radio is that it relies on *auditory perception*. The human voice is a more personal, direct and potentially more stimulating means of communication than the printed word, but this does not necessarily mean that radio is a more effective means of transmitting all kinds of communication to all kinds of persons. Studies done by the O.R.R., for example, indicate that people in the higher social and economic brackets more often prefer to read factual information rather than to hear it.

A third characteristic of radio is in part the outgrowth of the first two. Its accessibility, combined with its reliance on auditory perception, enables people to listen while carrying on a variety of other activities which do not necessarily interfere with their perception. But at the same time this quality of *non-interference* leaves the radio program liable to a low degree of attention. The listener may become so conditioned to it that he no longer hears it with any degree of acuity. This poses a problem for the investigator and necessitates a thorough probing of the seemingly factual statement: "Yes, I listened to that program" to determine the degree of concentration concealed behind it.⁹

A fourth characteristic of radio is that it *continues in time*. This means that a series of programs may become part of the daily or weekly habit patterns of the listeners, that *cumulative effects* can be built up over long or short periods. But it also

⁹ The problem of developing techniques to gauge degree of attention to radio programs is becoming a matter of great concern to television producers who want to know not only whether and how well a program was heard, but also whether and with what degree of attention it was seen.

means that it is liable to surfeit. It may be true that "if you hear a thing often enough you will come to believe it," but it is probably equally true that if you hear a thing too often you may not pay any attention to it at all after a time. Just where repetition ceases to be effective, just when saturation points are reached, is still a problem which has to be faced anew for each kind of program or message.

There are two other characteristics of radio which develop from its accessibility. A national network may reach into homes all over the country but if it does so it must *confine its appeal to a general one*. This national quality prevents it from appealing directly to local interests and experiences. Theoretically, of course, the potential audience is great enough for a nation-wide program to be beamed at special groups such as fishermen, students or stamp collectors and still reach a sizeable number of people. But since the aim of the networks is to reach as many people as possible at a given time, their specialized appeals are confined to large groups such as farmers or housewives who are known to constitute the majority of listeners at certain times of the day. Appeals to smaller groups are left for the local stations which cannot hope to compete with the networks on their own ground. The coming of frequency modulation might bring many changes in this respect.

The discussion of the nature of the problems met in the field of research and some of the techniques developed to meet them may be sufficient to bring the reader to two conclusions already evident to social scientists concerned with radio. First, there is a growing awareness of the necessity of systematizing the knowledge and experience accumulating in this field: a conviction that such self-conscious rigorization of procedures will be of value not only in the field of radio research alone, but to the science of communications in general (and perhaps even to other fields of social research). Secondly, as this formulation and formalization of procedures and problems proceeds, the sociologist and psychologist working in radio research become increasingly humble about what they do not know. But even at this comparatively early stage of their development, radio

research activities have already stimulated magazine and newspaper publishers to do more research than ever before, and it is possible that in the not too distant future not only will techniques and problems be exchanged between these two fields, but funds and research institutions may be merged for the greater benefit of both.

REFERENCES

1. Arnheim, Rudolph. "The World of the Daytime Serial." *Radio Research*, 1942-3. New York: Duell, Sloan and Pearce, 1944.
2. Beville, Hugh. *The Social Stratification of the Radio Audience*. No. 1 of Studies of the Social Psychology of Radio. New York: Office of Radio Research, 1944.
3. Cantril, Hadley, Herzog, Herta and Gaudet, Hazel. *The Invasion from Mars*. Princeton: Princeton University Press, 1940.
4. Curtis, Alberta. *Listeners Appraise a College Station*. New York: Federal Radio Education Committee, 1940.
5. Fiske, Marjorie. *A Sampling of Public Attitudes Toward the Proposed Subscription Radio Plan*. New York: Office of Radio Research, 1944.
6. Fiske, Marjorie. "How Consumers React to Radio Advertising by Retailers." *Sales Management*, LIII (1944), 92-98.
7. Fleiss, Marjorie. "Panel Interviewing as an Aid in Measuring the Effects of Advertising." *Journal of Applied Psychology*, XXIV (1940), 685-695.
8. Hallonquist, T. and Suchman, E. A. "Listening to the Listener" *Radio Research*, 1942-3. New York: Duell, Sloan and Pearce, 1944.
9. Herzog, Herta. "On Borrowed Experience." *Studies in Philosophy and Social Science*, IX (1941), 65-95.
10. Herzog, Herta. "What Do We Really Know About Daytime Serial Listeners?" *Radio Research*, 1942-3. New York: Duell, Sloan and Pearce, 1944.
11. Lazarsfeld, Paul F. *Radio and the Printed Page*. New York: Duell, Sloan and Pearce, 1940.
12. Lazarsfeld, Paul F. "The Daily Newspaper and its Competitors." *The Annals of the American Academy of Political and Social Science*. Philadelphia, 1942.
13. Lazarsfeld, Paul F., Berelson, Bernard and Gaudet, Hazel. *The People's Choice*. New York: Duell, Sloan and Pearce, 1944.
14. Lazarsfeld, Paul F. and Fiske, Marjorie. "The Panel as a New Tool for Social Research." *Public Opinion Quarterly*, II (1938), 596-612.
15. Lazarsfeld, Paul F. and Kendall, Patricia. "The Listener Talks

- Back." *Radio in Health Education*. New York: Columbia University Press, 1945.
16. Lazarsfeld, Paul F. "Audience Building in Educational Broadcasting." *Journal of Educational Sociology*, XIV (1941), 533-541.
 17. McCandless, Boyd. "A Study of Non-Listeners." *Radio Research, 1942-3*. New York: Duell, Sloan and Pearce, 1944.
 18. Meyrowitz, Alvin and Fiske, Marjorie "The Relative Preference of Low Income Groups for Small Stations." *Journal of Applied Psychology*, XXIII (1939), 158-162.
 19. Robinson, William S. "Radio Comes to the Farmer." *Radio Research, 1941*. New York: Duell, Sloan and Pearce, 1941.

THE DUTIES OF CIVIL SERVICE EXAMINERS AND TEST TECHNICIANS

THE following check list is one of a series being prepared under the auspices of the Society for Public Administration to depict what public personnel workers do. It is a modification of a list originally prepared by Dr. John M. Piffner for the Committee on Education of the Society. It is presented here in the hope that readers of EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT will offer their constructive criticisms before the list is included in the final published report. Comments will be welcomed by the chairman of the Committee, Mr. Edgar W. Lancaster, Office of the Secretary of War, Room 4E 978, Pentagon, Washington 25, D. C.

It should be noted that all of the duties listed below are not ordinarily performed by any one person. The list is intended to cover the work which may be done by all workers engaged in merit system examining. A list of the main areas of activity is given first, followed by the more detailed outline.

Main Topics

- I. Plan examinations.¹
- II. Construct tests.
- III. Supervise the scoring of tests.
- IV. Evaluate training and experience.
- V. Supervise the administration of tests.
- VI. Supervise the conduct and scoring of competitive oral interviews.
- VII. Establish registers of eligibles.
- VIII. Serve as consultant for the service rating program.

¹The term "examination" is used in a broad sense to indicate the entire procedure by which a person's qualifications are evaluated with respect to a position or series of positions. An examination may include a test or tests, an oral interview, or an appraisal of training and experience, either singly or in any combination.

IX. Participate in establishing classification specifications.

X. Conduct research on examinations.

Detailed Outline

- I. Plan examinations.
 - A. Assemble data concerning the duties of the positions for which examinations are to be conducted.
 1. Consult class specifications.
 2. Secure additional job descriptions from operating officials.
 3. Search for additional job standards.
 4. Interview supervisors, foremen, and workers.
 5. Observe actual work being done and do some of the work, if practicable.
 6. Read laws, regulations, directives, manuals and books which have a bearing on the job.
 7. Prepare summary showing
 - a) Duties.
 - b) Knowledge necessary.
 - c) Skills used.
 - d) Personal characteristics required.
 - B. In the light of information obtained in Step I-A, determine the minimum requirements for taking the examination, if any, within the limits set by legislation.
 - C. Outline the nature of the examination considered to be most suitable for selecting qualified workers. Within the limits set by legislation, and usually after consultation with experts in the field, determine whether it shall include an evaluation of training and experience, a competitive oral interview, and written or performance tests, giving due consideration to any pertinent experimental data which may be available.
 - D. Determine weights to be assigned the major parts of the examination, taking all pertinent information into account.

E. Prepare the copy for an announcement of the examination for the printers, or pass on the requisite information to those in the organization charged with the responsibility of issuing announcements. Include information on the following topics:

1. Title of position.
2. Grade and salary of position.
3. Location of employment.
4. Hours of work.
5. Detailed explanation of education and experience requirements.
6. General information.
 - a) Whether written tests, performance tests, and competitive oral interviews will be given.
 - b) Provisions of regulations regarding qualifications statements.
 - c) General eligibility requirements in regard to citizenship, veterans' preference, etc.
 - d) Information as to how and where application should be made.
 - e) Weights to be given the various parts of the examination.

II. Construct tests according to plan.

A. Construct written tests.

1. Select promising existing test items from file, noting statistical history resulting from item analysis.
2. Edit old test items to suit current need.
3. Write new test items, observing the best psychological and psychometric techniques and procedures. In the case of achievement tests:
 - a) Read widely in subject-matter field.
 - b) Collect background material.
 - c) Confer with other personnel technicians.
 - d) Ask subject matter experts to submit test material.
 - e) Train subject matter experts in test construction.

4. Check on the final content of the test.
 - a) Check items for possible defects such as faulty phrasing or the presence of specific determiners.
 - b) See that items are of appropriate difficulty.
 - c) In the case of achievement tests, make sure that concepts within the required areas are adequately sampled and that each area is appropriately represented.
 - d) Check to see that repetition of concepts is avoided and that the content of one question does not help in answering other questions.
 - e) Check achievement test items with operating officials and subject matter experts.
 5. Assemble the test in final form with written directions, instructions for administration, answer keys, and directions for scoring.
 - B. Develop performance tests, when called for, on the basis of the study made in step I-A, giving due consideration to ways of measuring the process of performance as well as the final product.
- III. Supervise the administration of tests.
- A. Arrange for the use of suitable room or rooms.
 - B. Arrange for enough qualified proctors.
 - C. Arrange to have tests, pencils and necessary apparatus ready, with precaution taken to prevent tests from being inspected before they are given.
 - D. Train proctors.
 - E. Administer tests.
- IV. Supervise the scoring of tests.
- A. Supervise the preparation of scoring keys for objective-type questions.
 - B. Plan the scoring procedure for questions which are not objective, taking steps to insure reliability and uniformity of scoring.
 - C. Develop the scoring procedure for performance tests, if used.
 - D. Train scorers.

- E. Determine weights to be given parts of the test, taking all pertinent information into account.
 - F. Obtain total scores in accordance with the plans developed for the examination.
 - G. Transmute total scores to a basis appropriate for combining with any other measures of qualifications which may be used, with due consideration for any legal requirements which may exist.
- V. Evaluate training and experience.
- A. Determine the amount of credit, if any, to be given various types of experience, obtaining the advice of experts in the field and making use of validation studies, where possible.
 - B. Determine the amount of credit, if any, to be given various types of training on the basis of the advice of experts in the field and the conduct of validation studies, where possible.
 - C. Develop a schedule to be used in evaluating training and experience in accordance with the credit system developed.
 - D. When the rating is not done by a rating technician or subject matter expert, train clerks in the use of the form for evaluating training and experience and supervise them in applying the procedure.
 - E. Decide special questions which arise in connection with evaluating training and experience.
 - F. Transmute raw scores on training and experience to a base appropriate for combination with other measures, with due consideration for any legal specifications which may exist.
- VI. Plan and supervise the conduct and scoring of competitive oral interviews when such interviews are employed by the merit system.
- A. Determine characteristics to be measured by the oral interview as distinguished from those covered in other parts of the examination.
 - B. Develop instructions for interviewers.
 - C. Develop rating scales for factors to be observed in the oral examination.

- D. Outline desirable qualifications for members of interviewing boards.
 - E. Train members of interviewing boards.
 - F. Develop and apply a method of obtaining scores based on competitive oral interviews.
 - 1. Obtain a score for each interviewer's report.
 - 2. Combine the scores from all interviewers for each applicant, taking necessary steps to insure that ratings are given equal weight.
 - 3. Transmute the scores to a standard scale appropriate for combining with other measures, with due consideration for legal requirements.
- VII. Establish registers of eligibles.
- A. If weights were not previously established, determine the weights for the component parts of the examination, taking all pertinent information into account.
 - B. Combine component parts of the examination in order to give established weights to the various parts.
 - C. Set a passing point and transmute the scores to the standard grading system in use.
 - D. Adjust final scores for special preference groups, if any.
 - E. Supervise the preparation of the register with the eligibles placed in the order of final score.
- VIII. Serve as consultant for service rating program
- A. Assist in developing the rating schedules to be used and the method of scoring them.
 - B. Participate in the analysis of the results from service ratings.
 - C. Assist in handling problems arising from the use of service ratings.
- IX. Participate in establishing classification specifications.
- A. Conduct studies on the relation of minimum requirements of education and experience to effectiveness of job performance.
 - B. Consult with classification analysts on the analysis

of required knowledges, skills, and abilities in meaningful, measurable psychological terms.

- C. On occasion, apply measurement techniques to the evaluation of factors determining the allocation of positions to specific grades.
- X. Conduct research on examinations.
 - A Conduct validation studies, preferably before tests are used, if the need for maintaining the confidential nature of the tests allows.
 - 1. Give to a group or groups of persons comparable to those to whom the test will be applied.
 - 2. Determine criteria against which the tests should be validated and develop reliable measures of the criteria, when possible
 - 3. Obtain measures of the relation between the tests and the criteria. Make a study of the validity of the test items.
 - 4. Revise the test or scoring procedure in the light of studies of validity.
 - 5. Continue validation studies through investigating the relation of test scores to subsequent performance on the job, with a view to discovering principles to be used in constructing valid tests for similar positions in the future.
 - B. Check on the reliability of tests and conduct studies leading to the construction of tests of adequate reliability.
 - 1. Obtain estimates of reliability of individual tests.
 - 2. Conduct item analyses to determine internal consistency of sets of items.
 - 3. Conduct studies designed to help in eliminating items which tend to lower the reliability of tests.
 - C. Make item analyses of tests used and set up a system for maintaining records of the performance of items for various groups and examinations.
 - D. Conduct research on miscellaneous measurement problems.

THE ARRANGEMENT OF CHOICES IN MULTIPLE CHOICE QUESTIONS AND A SCHEME FOR RANDOMIZING CHOICES

CHARLES I. MOSIER AND HELEN G. PRICE

State Technical Advisory Service, Social Security Board

IN the construction of objective test items in multiple-choice and allied forms, the arrangement in order of correct answer and choices is often a troublesome chore. Some test writers prefer certain positions for the correct choice, finding that items having the correct answer in a position other than as the first or last choice increases the difficulty of the item;¹ most, however, resort to one device or another to secure a systematic, truly random arrangement. When the location of the correct choice in the series of choices is left to the whim of the test constructor, personal position-preferences will almost inevitably result in a preponderance of correct answers falling in one position. Moreover, since distracters tend to be written in order of plausibility, with the last distracter often written as a desperate final effort, a randomization process should extend beyond the correct choice to the incorrect ones as well. The present paper presents a "randomizer" for five-choice items, a discussion of its use, and a simple method by which other similar aids may be constructed. Some of the situations in which it should not be used are also considered.

In writing multiple-choice items, we at the State Technical Advisory Service follow certain mechanics designed to simplify the writing process and to provide needed controls on quality. After the premise or the question has been formulated, the intended answer is always written first, with the incorrect choices following. This practice of having the intended answer written

¹McNamara, W. J. and Weitzman, E. "The Effect of Choice Placement on the Difficulty of Multiple Choice Questions" *Journal of Educational Psychology*, XXXVI (1945), 103-113

as the first choice has a number of advantages. It insures that a correct choice is included and that the item writer, in his zeal to prepare plausible distracters, does not end with five plausible choices—all wrong. (Such things have happened.) Moreover, in case of any later doubt that the answer indicated on the scoring key may not be the one intended, a quick check against the original draft will remove any question about the writer's intent. Editorial review and checking are also facilitated.

After the items are reviewed for authenticity and edited for grammatical construction and technical form, the alternative answers are assigned their final, random order by use of the table attached. The table, constructed for five-choice questions, shows the 120 permutations of the numbers one through five. In preparing the table the permutations were written in systematic, cyclic order and each permutation was assigned a sequence number from one through 120. Each permutation was then assigned, as its final position in the table, the order in which its sequence number occurred among the last three digits of a nine-place table of logarithms.

The use of this table has these advantages over other systems of randomization: (1) only the numbers actually used need be considered; (2) there are no repetitions or omissions of choice numbers for any item; and (3) the order of all five choices is given simultaneously. In applying the table, the item writer assigns to each successive item the choice patterns in the order in which they occur, beginning each new group of items where the previous one left off; every choice pattern is thus used once before any pattern is repeated.

The present table can be used for three- and four-choice items as well. Such use, however, loses the principal advantages enumerated above and it would seem far simpler to use the same procedures in constructing similar "randomizers" for those types of items.

For certain types of items the choices should not be randomized. When the choices represent selections from a meaningful, ordered series, e.g., dates or magnitudes, it is far less confusing to the candidate if they are arranged in their natural order. Even where the order of choices is fixed, e.g., by a series of

dates, the randomizing table can be used to advantage in locating the correct choice objectively, thus determining the number of dates the item writer can select which should precede and follow the correct one. By using the table the item writer can select the choice pattern, assign to the correct choice the first number in the pattern, and distribute the distracters around the choice. Thus, if the choice pattern were 2 4 3 5 1 and the question were:

"The year in which the Pilgrims landed at Plymouth Rock is": the intended answer would be choice 2 and the completed arrangement of choices might be:

(1) 1607; (2) 1620; (3) 1628; (4) 1636; (5) 1776.

A predetermined choice pattern should be used, of course, only when the incorrect choices are selected because of their association with the question asked; it is more important to have effective distracters than to follow a prescribed order.

Another situation in which the choices should not follow a randomized pattern is that in which the choices include any of the numbers one to five as answers. In these items, the number of the correct choice should be the same as the choice itself; thus:

"The reciprocal of .25 is: (1) one (2) two (3) three (4) four (5) five."

It is sufficiently confusing to the candidate to have to answer the question without having to remember, as he might if the choices were randomized, "The answer to the question is four but 'four' is choice number 3 and it is not the answer, four, but the number of the answer, 3, which I must mark in the answer booklet." The problem posed by this type of answer can, of course, be met by shifting from the designation of choices by number to the designation by letter. The use of letters, however, has disadvantages and the other solution seems preferable.

There is another situation in which it seems desirable to modify the random order of choices. The discussion of this situation is presented here as an hypothesis partially borne out by observation rather than as one verified by evidence. When the choices presented include a best answer and another which,

although not the best, is very nearly as good, the writers have observed that the item will very frequently have a negative item-test coefficient if the nearly-as-good choice precedes the best answer; the coefficient is far more likely to be positive if the order of these two choices is reversed. Apparently the high-scoring candidates read until they come to the not-quite-so-good choice, recognize it as an acceptable answer, give it, and turn immediately to the next question. The lower-scoring students, unable to find an answer which their knowledge will permit them to identify as correct by recognition, make a careful comparison among the alternatives and have a greater probability of success than the high-scoring group. Reversing the order of the two choices makes the item easier, but tends to correct its negative relation with total score. Whether such a change will have the desired effect of increasing the discriminatory power of a particular item must, however, be weighed very carefully in the light of the choices in question, the function to be served by the item, and the group of candidates for which it is intended.

NUMBERS FOR RANDOMIZING 5-CHOICE ITEMS

35124	13425	12453	24531	54231	42153
34215	43512	54132	32541	32154	25143
12354	43251	15324	52314	21453	24135
25134	51243	25314	42315	54321	52143
14523	13254	15423	51342	12543	34152
24513	43152	54312	42531	23451	52134
52431	31245	13524	51234	21534	15342
31452	42513	12534	14532	35412	23514
54213	25341	43521	41325	53421	34251
53412	34521	21435	32514	51432	32145
43125	13542	35241	41352	52341	51324
23541	13245	45132	31542	54123	43215
45231	41235	24351	31425	23415	41523
21354	14235	23145	52413	35214	24153
15243	53124	21543	31524	25431	41532
42135	21345	25413	12435	35142	34125
31254	45213	32415	12345	51423	23154
14253	45321	53214	35421	14352	24315
53142	13452	42351	45123	32451	45312
15432	41253	14325	53241	15234	34512

THE RATIONALE OF TEMPERAMENT TESTING

DONCASTER G HUMM

Personnel Service, Los Angeles, California

Temperament

TEMPERAMENT, according to the dictionary, has to do with internal constitution. It also has to do with the peculiar physical and mental characteristics that influence an individual's disposition, or the character of mind or mental reactions having to do with his behavior. Hence temperament may be considered as the pattern or complex of tendencies which determines an individual's behavior. As such, it is made up of traits or tendencies to respond in a consistent manner whenever a given type of situation arises. Each individual has an abundance of traits arising out of the interaction of his original nature and his environment; some are chiefly the result of hereditary forces; others, of non-hereditary forces; and still others, of mixed forces. Some inhibit the effect of others while some tend to reinforce others.

Temperament may be analyzed in many different fashions depending upon the psychological attack of the author. Thus, Freud analyzes it into the following three tendencies to react: (1) toward self-preservation, (2) toward race-preservation—or sex, (3) toward gregariousness. Jung, with his emphasis upon attitudes, divides temperament into two great types: the introverts and extroverts. This oversimplification, however, should not be accepted as Jung's final analysis, since he subdivides introverts and extroverts into several different subclasses and also considers the ambivert, the individual who has characteristics that are both extroverted and introverted.

One of the most useful analyses of temperament is Rosanoff's (1). This analysis has the merit of reflecting the practical experience of many prominent students of personality including

Leri, Birnbaum, Kraepelin, Spratling, Davenport, Dostoyevski, and Flaubert. Rosanoff's function has been that of editor plus that of integrator plus that of contributor. He has taken the practical experience of these men, added his concept of the control and directive component of temperament, and integrated all of this psychiatric experience into a comprehensive analysis of the field. It should be noted that this analysis is chiefly the result of the experience, observations, and research of psychiatrists dating back as far as the 1870's. We used it as the basis of the *Humm-Wadsworth Temperament Scale* because it had already demonstrated its value to us by explaining problems of behavior in clinical and industrial use.

The Characteristics of a Good Temperament Test

The first characteristic of a good temperament test is that it is based upon a comprehensive and valid analysis of temperament. As such, a temperament test may be based on any of the three analyses we have mentioned or on any analysis which is sufficiently comprehensive to cover the reaction pattern of an individual.

The second characteristic of a good temperament test is adequate standardization. Adequate standardization starts with a good sampling of the temperamental traits which must be explored in a comprehensive analysis. These samples ordinarily consist of questions or items which may have a variety of forms. They may be multiple-choice, false-or-true, or any of the types of items which will bring to light the basic traits possessed by the test subjects. Each of these test items must be subjected to a careful item analysis to determine whether or not it actually elicits the information desired.

It is important in constructing a temperament test to take account of the purpose for which the test is to be made since the attitude with which the individual responds to the test will influence his answers. If the test is to be used for clinical purposes, the control subjects on whom the test is standardized should have the atmosphere of the clinic in which to respond to the items or answer the questions. If the test is to be an industrial test, the atmosphere which pervades the testing of

applicants or workers for industry must also be present when the control subjects respond to the questions. However, it has been possible in our experience to develop measures by which compensations for subjects' attitudes can be made so as to increase the usefulness of the test.

Thus, in the standardization of the *Humm-Wadsworth Temperament Scale* (2) two measures of the subject's attitude were studied: the No Count and the Profile Count. These revealed a tendency on the part of the subject to report his temperamental tendencies in an atypical manner or to overreport them or to underreport them. In the *Manual of Directions* (3) which accompanies the *Human-Wadsworth Temperament Scale*, statistical compensations are reported to make it possible to consider overreported and underreported Scales as though they had been typically reported. This subject is further considered in the *Manual of Interpretation* (4). There has also been provided a *Nomograph* (5) to make it possible to make these compensations more easily. A simple explanation of the compensations is also included in *Personnel Evaluation Method* (6).

After the items (or questions) of the test have been evaluated the next step is the construction of norms. In this regard, temperament tests are very different from other types of tests such as interest inventories, intelligence tests, skill tests, and the like. It is very difficult, if not impossible, to construct a temperament test with only a single set of norms. This arises out of the peculiar nature of temperament.

In intelligence-test construction, the objective is to find a measure by which the subject may be compared with the whole population, so the procedure is to standardize the projected test on a control group which represents as nearly as possible a cross-section of the population. In temperament-test construction many of the tendencies we are trying to measure are very difficult, and perhaps impossible, to identify in average well-controlled individuals by any means available before the test is made. Moreover, some of these tendencies occur with less frequency than do others; in any survey of a cross-section of the population, they appear as statistically unimportant, but in the

individuals in whom they are strong they have the utmost importance.

In standardizing the *Humm-Wadsworth Temperament Scale* we hit upon the device of using control subjects, selected by case study, who represented extreme examples of the tendencies we wished to measure either by possessing the tendencies to a high degree or by not possessing them at all. The following tabulation will illustrate our use of such control groups.

Control Groups Used in Standardizing the Humm-Wadsworth Temperament Scale

<i>Components</i>	<i>Plus Groups</i>	<i>Minus Groups</i>
"Normal"	Strongly "Normal" Subjects	State Hospital Patients
Hysteroid	Habitual Criminals	Self-Sacrificing Persons
Manic	Excitable, Emotional Subjects	Subjects lacking Manic Traits
Depressive	Strongly Depressive Subjects	Subjects lacking Depressive Traits
Autistic	Shy, Seclusive Subjects	Subjects lacking Autistic Traits
Paranoid	Aggressive, Opinionated Subjects	Subjects lacking Paranoid Traits
Epileptoid	Subjects given to Epileptoid Tendencies	Subjects lacking Epileptoid Traits

The Scale as developed in this way then gave us a description of the individual's disposition, a measure of his mental health, and a comparison of his tendencies to react with those of other typical groups. Thus, a temperament test describes the disposition of the subject, estimates his powers of self-mastery and self-control, and compares his reaction pattern with the reaction pattern of other subjects of known characteristics. Having provided for such a picture of test subjects by comparison of their scores with those of specially selected subjects we proceeded to learn the meaning of our scores in terms of the average, that is the general, population. This was accomplished by giving the test to a large group of adults (all of the employees, from president to unskilled laborers, of a company which had not previously used tests). This group is probably not a perfect cross-section of the whole population but we have good reason to believe that it is a satisfactory sampling.

The distribution of scores afforded by this survey gave us information as to the average strength in well-adjusted adults

of the tendencies measured. We found, for example, that tendencies to be sociable, cheerful, active, and emotionally responsive to the environment are relatively common, while tendencies to be conceited, suspicious of the motives of others and stubbornly fixed in one's opinions, are fairly rare.

We are frequently asked by personnel men and by students why we cannot provide for an overall score which might be taken as a general measure of good or poor temperament. We do not do this because an important consideration in the use of temperament tests is the identification of *patterns* of behavior tendency. Thus, such a test really is a battery of tests rather than a simple measure. The interrelationships among the various measures included in the battery are quite certainly more important than the strength of the individual components of temperament considered separately. This consideration of temperamental patterns or syndromes enormously complicates both the construction and interpretation of temperament tests. As a result, the problem of making a temperament test becomes an expensive and time-consuming project, and the problem of interpreting the completed test is one which requires the acquisition of special skills. I suspect that all types of tests would gain in usefulness if we would pay more attention to the specialized problems of interpretation each type presents.

I have mentioned the problem arising out of the attitude with which the subject approaches the test situation. In intelligence testing and skill testing it may be assumed that most subjects will do the best they can—except, perhaps, for such situations as those in which a criminal might feign feeble-mindedness or a soldier try to conceal a skill which would lead to an undesired assignment.

In temperament testing, however, all sorts of complexities affect the subject's responses. There are, of course, no "right" answers or "wrong" answers. Each answer will be true for some subjects and untrue for others. It is often supposed that the expected or favorable answer would be easily recognized and would be selected by all or most applicants for jobs, but this does not happen. Some subjects seem to be more suggestible than others, either positively or negatively; some seem to lean

over backwards to claim unfavorable traits; others seem to deny the possession of even desirable characteristics. A successful temperament scale must include in its scoring and interpreting procedures means of taking account of these tendencies. We have found that certain relationships among the scores reveal the effect of these attitudes, and compensation for such attitudes can be made.

Use of Temperament Tests

As noted previously, a good temperament test may be used for a prediction of behavior since it will reveal the status of the subject's mental health, it will describe his disposition, and it will compare his behavior tendencies with others. However, a temperament test cannot be used as a prediction of behavior unless the situation in which this behavior is to occur is carefully taken into account and unless the other factors of personality, aside from temperament, are also taken into account. This follows from the fact that a temperament test reports tendencies—tendencies which are operative only in the presence of trigger situations. Thus, a temperament test should be so constituted as to report the probable behavior of an individual when he is free from undue strains and also report his probable behavior when he overcompensates for strains.

All this makes the estimate of the situation and the estimate of other factors influencing behavior, as summarized in the estimate of probable strains, an important consideration in the use of temperament tests.

The situation in which an individual is placed may or may not be of such a nature as to be conducive to a tranquil, acceptable adjustment of the individual. If it may be taken for granted that the individual is in a compatible, sympathetic, and kindly atmosphere, it may be taken for granted that strain in such a situation will be reduced to a minimum. If, however, the situation has anything in it which is likely to put the individual on the defensive or likely to give rise to contention or other forms of unpleasantness, it can be predicted that the individual will undergo strain. Thus, it follows that the findings of a temperament test alone are not sufficient to predict how

an individual will respond in any given special situation. For example, it is not possible to predict from the findings of a temperament test how well a student will adjust to college unless it is also possible to predict how well the student will like the college atmosphere, how suitable his course will be for his aptitudes and interests, and how well he will be received. Similarly, it is not possible to predict how well a worker will get along with a group of workers unless it is possible to predict how well he will like the group, including his boss, how well the group will like him and get along with him, and how well the job will fit him.

There are several factors in the constitution of the individual, aside from his temperament, that have an influence on his behavior. Some of these are in the field of aptitude. For example, if an individual is placed in a business situation where his intelligence is not adequate, one must expect an undue strain to result. If he is placed where his intelligence is so superior to the job that it is very incompletely utilized, one must expect another sort of strain—that of boredom. This reaction is also to be expected with reference to skill. A highly skilled worker placed in a job which makes demands for mediocre skill is likely to become dissatisfied and get into mischief. A worker who is placed in a position which requires more skill than he possesses is likely to become discouraged or defensive or in some other way to compensate for his feelings of inadequacy.

Health is also an important consideration in estimating strain. For example, a man of super-abundant energy with considerable pressure of activity cannot be tied down to an inactive job without the expectation of some over-compensation on his part. Likewise, an individual who is struggling in a job beyond his strength is likely to suffer, not only with respect to his physical health, but also with respect to his mental health.

It follows that the prediction of behavior can be accomplished by the use of a temperament test if the findings of that test are supplemented by findings of other sorts—probably including non-test data as well as test data—to predict the amount of strain the individual may be expected to endure in the situation or situations under consideration.

Temperament Testing in Clinical Practice

Valid temperament tests are useful in studies of individuals for vocational guidance, educational guidance, and problems of social and marital adjustment. In many instances a good temperament test will indicate whether or not a personal problem will be complicated by a poor state of mental health or by an insufficiency of self-control or self-mastery to direct dynamic temperamental qualities. It should be the practice of a psychologist, however, to refer problems of mental health to psychiatrists for examination. Whenever there is any question of psychosis, psychoneurosis, or a psychopathic state, it is necessary to consider not only the behavior of the individual but also the nature of the handicap or disablement. This makes it essential to secure a medical diagnosis as well as a psychological diagnosis. Psychiatrists only are equipped professionally to consider both these phases.

A painstakingly thorough study of personality is required in the consideration of the readjustment of the individual. It seems reasonable that the minimum points to be covered are the following: (1) family history, at least as far back as the grandparents, in which noteworthy achievements and handicaps are taken into account; (2) personal history from conception, including childhood, adolescence and adulthood; (3) a particularized history of difficulties in making adjustment—especially the failures in school and social and job adjustments; (4) a physical examination by a competent physician; (5) a preliminary mental examination by means of a valid temperament test followed by a verification of the results in a personal interview or by psychiatric examination; (6) an interest examination by a standardized interest inventory; (7) examination of skills and aptitudes by competent tests; (8) examinations by intelligence tests—preferably by an individual intelligence test; (9) the analysis of all of the data obtained in steps one to eight and a report to the subject; (10) a written summary and report to the subject.

The use of such a procedure is very likely to be effective in substantiating and explaining the individual tests by the results of the tests in other fields. Such a procedure is more than

likely to bring to light the extent to which the individual has undergone strain, the character of that strain, and its probable effect upon his temperamental integration.

Temperament Tests in Industry

Temperament tests in industry are valuable to supplement aptitude tests and data obtained by non-testing procedures. Aptitude tests tend to reveal to the technician what the individual *can* do; temperament tests, what he *will* do; and interest tests, what he *likes* to do. The integration of testing methods and non-testing methods in a routinized procedure is very likely to prove more effective than the use of either test procedures or non-test procedures alone. A good industrial appraisal program probably should include the following:

(1) A standardized application form or job-specification-and-qualification sheet. This form should contain spaces for background, training, experience, job titles, and job duties.

(2) Intelligence tests; if group tests are used, at least two should be included. When possible, one of these should be a timed test and one an untimed test. Some individuals do not respond well to timed tests.

(3) A temperament test; this test determines the individual's self-mastery and self-control, the strength of his temperamental characteristics, and his behavior tendency pattern.

(4) An interest inventory; this measure determines whether or not the individual's interests are such as to make him contented in the type of work being considered.

(5) Skill or aptitude tests; skill tests are to be preferred where the individual is already trained for the contemplated job. Aptitude tests are to be preferred where the individual is a trainee.

(6) Physical examination by the company physician.

(7) Summary of all of the data considered in the foregoing six procedures, a listing of assets and liabilities with regard to the job, and a statement of job risk.

Such a set of procedures as this can be so routinized as to take less than three hours' time. The fact that many of the tests are

group tests makes it possible to test many individuals simultaneously. However, the most important feature of such a set of procedures is its thoroughness. After all of these points have been covered, it is possible to have such an understanding of the potentialities of the worker as can be used for selection, placement, counseling, supervision, and readjustment.

Summary

Temperament is one aspect of personality, but not the whole of personality. It is the pattern or complex of tendencies which express themselves in behavior in the presence of trigger situations.

The measurement of temperament requires: (1) a valid and comprehensive analysis of temperament as a base of departure, (2) items or questions which adequately sample the field of temperament; (3) adequate item analysis; (4) norms which afford a description of temperament and comparisons with the population; (5) provision for dealing with atypical response attitudes.

Temperament tests may be used for the prediction of behavior when other pertinent facts are known; that is, when the environmental strain can be estimated. They cannot be used for such prediction unless environmental strains are considered. (Incidentally, environmental strain cannot be taken as a constant. Even the conditions of combat represent for some men a challenge or opportunity or release, while for others they represent only danger or sorrow or frustration.)

Temperament tests, properly used, can be valuable aids to the clinical and industrial psychologist for the information they give with respect to mental health, temperamental integration, strength of various temperamental characteristics, and behavior patterns.

REFERENCES

1. Rosanoff, A. J. *Manual of Psychiatry*. 6th ed. New York: Wiley, 1927.
2. Humm, D. G. and Wadsworth, G. W. "The Humm-Wadsworth Temperament Scale." *American Journal of Psychiatry*, XCII (1935), 163-200.

3. Humm, D. G. and Wadsworth, G. W. *Manual of Directions* for the Humm-Wadsworth Temperament Scale. Los Angeles: D. G. Humm, 1942 ed.
4. Humm, D. G. and Wadsworth, G. W. *Manual of Interpretation* for the Humm-Wadsworth Temperament Scale. Los Angeles: D. G. Humm, 1943.
5. Humm, D. G. *Nomograph* for the Humm-Wadsworth Temperament Scale. Los Angeles: D. G. Humm, 1942 (revised 1944).
6. Humm, D. G. *Personnel Evaluation Method*. Los Angeles: D. G. Humm, 1945.

MECHANICAL ABILITY, ITS NATURE AND MEASUREMENT. II. MANUAL DEXTERITY

J. R. WITTENBORN

Yale University

Introduction

The factor analyses of test samples employed in studies of mechanical and motor abilities by Harrell (3) and Wittenborn (6) have shown that the variables may be classified on the basis of their interrelationships. These classifications, or factors, offer a functional basis for the definition of abilities. In the analyses of mechanical ability these factors appear to be of two types and for the sake of simple designation may be given the superficially descriptive labels of "mental" and "motor." These rubrics are, of course, not explanatory, but they are appropriate insofar as the variables contributing to the "mental" abilities (scholastic, spatial visualizing, and perceptual) are considered to be independent of the exact mode of expression. Tests contributing to the "motor" abilities (dexterity, repetitive movement, and steadiness) appear to be peculiarly dependent upon the quality of muscular performance.

The present paper is concerned chiefly with "motor" abilities, particularly those which may be called "manual." It is based primarily on data from the Experiment Proper of the Minnesota Mechanical Ability program of research (4). The Minnesota program had two aims: one was to predict "mechanical ability" for a group of junior high-school boys in shop courses; the other was to understand the general nature of mechanical ability, something about its origins, and the conditions for its development. As a part of the Minnesota study of the nature of mechanical ability, the following variables from the Experiment Proper were intercorrelated:

- | | |
|---|---|
| 1. Age | 26. Father's mechanical operations |
| 2. Otis, I.Q. | 27. Tools owned by son |
| 3. Packing Blocks | 28. Tools owned by father |
| 4. Card Sorting | 29. Things done questionnaire |
| 5. Minn. Spatial Relations | 30. Mechanical occupations preferences |
| 6. Paper Form Board | 31. Academic preferences |
| 7. Stenquist Picture I | 32. Interest Analysis Blank (old) |
| 8. Stenquist Picture II | 33. Gymnasium ranks |
| 9. Minn. Assembly | 34. Academic grades |
| 10. 100-yard dash | 35. Garfield's Agility battery |
| 11. Back dynamometer | 36. Minn. Agility battery |
| 12. Right-hand dynamometer | 37. Interest Analysis Blank (new) |
| 13. Steadiness | 38. Shop operations quantity-quality criterion |
| 14. Left-hand dynamometer | 39. Education of father |
| 15. 25-yard hop | 40. Education of mother |
| 16. Spirometer | 41. Mechanical ability rating of father's occupations |
| 17. Broad jump | 42. Mechanical ability rating of other ancestors' occupations |
| 18. Height | 43. Barr scale ratings of father's occupations |
| 19. Weight | 44. Barr scale ratings of other ancestors' occupations |
| 20. Shop operations quality criterion | 45. Otis, mental age |
| 21. Shop operations information criterion | |
| 22. Cultural status | |
| 23. Literary interests | |
| 24. Recreational interests | |
| 25. Son's mechanical operations | |

An examination of the intercorrelations revealed that numerous variables, such as numbers 22, 23, 24, 27, 28, 29, and 30, bore no important relationships with other variables. Certain other variables, such as 35 and 36, 39 and 40, 41 and 42, 43 and 44, tended to form independent couplets and as a consequence were of no general interest. Mental age and other variables relating to academic status were of no interest in the present study, and the sole steadiness test, variable 13, showed no important relationship with any of the other variables. Certain of the variables, however, showed significant interrelationships, and their nature suggested that further scrutiny might afford additional insight into the nature and organization of mechanical ability. These promising variables and their intercorrelations are presented in Table 1.

An Analysis of the Minnesota Data

The 16 variables for which intercorrelations are shown in Table 1 were selected with certain expectations. It was believed, for example, that the pattern of their intercorrelations might confirm the tendency for measures involving a high degree of manual dexterity to form an independent functional classification (6). It was expected, moreover, that an analysis

TABLE I
Intercorrelations of 16 Selected Variables (N=100)

	Packing Blocks	Card Sorting	Minn Spatial Relations	Paper Form Board	Stenquist Picture I	Stenquist Picture II	Minn Assembly	Back Dynamometer	Right-Hand Dynamometer	Left-Hand Dynamometer	Spirometer	Height	Weight	Shop Operations Quality Criterion	Son's Mech. Operations	Interest Analysis Blank (new)
	3	4	5	6	7	8	9	11	12	14	16	18	19	20	25	37
3																
4	.52															
5	.34	.23														
6	.14	.14	.63													
7	.18	.10	.42	.37												
8	.21	.24	.39	.30	.54											
9	.30	.13	.56	.49	.46	.40										
11	.09	.13	.11	-.01	.25	.11	.15									
12	-.03	-.04	-.05	-.10	.08	-.12	.04	.66								
14	-.06	-.06	-.09	-.09	.15	-.10	.06	.70	.84							
16	-.07	.03	-.01	.03	.16	-.02	.04	.54	.50	.60						
18	-.02	-.09	.01	-.08	.16	-.11	-.01	.48	.58	.60	.72					
19	-.09	-.11	-.01	-.05	.18	-.04	.04	.59	.67	.68	.74	.78				
20	.26	.19	.53	.52	.24	.31	.55	.04	.05	.02	.03	-.09	.02			
25	.00	-.12	.22	.24	.24	.19	.40	.11	.15	.09	.04	.09	.10	.30		
37	.12	.09	.46	.39	.32	.28	.42	.08	.09	.04	.03	-.01	.03	.64	.30	

of the selected intercorrelations would contribute to our understanding of the general nature of the dexterity factor. It seemed to be particularly desirable to know to what degree measures of strength and physical development contribute to an ability such as manual dexterity.

The intercorrelations were subjected to a centroid analysis and four factors were extracted. No residual significantly greater than zero remained. When the centroid matrix, Table 2, is postmultiplied by the transformation matrix, Table 3, the orthogonal rotated factor matrix, Table 4, is produced.

Although an orthogonal solution is given to the present problem, it is apparent that Factors I and II are not truly independent. The variables which cluster together to form Factor II have higher loadings on Factor I than on Factor II. It is apparent, therefore, that presentation of Factor II as a factor independent of Factor I is not in strict conformance with the

TABLE 2
Centroid Matrix

	I	II	III	IV	h^2
3. Packing Blocks31	.33	.39	.36	.49
4. Card Sorting24	.26	.45	.42	.50
5. Minn. Spatial Relations56	.52	-.12	.12	.61
6. Paper Form Board46	.48	-.25	.09	.51
7. Stenquist Picture I56	.23	.12	-.21	.43
8. Stenquist Picture II40	.43	.23	-.15	.42
9. Minn. Assembly59	.45	-.10	-.16	.58
11. Back dynamometer61	-.47	.27	-.09	.67
12. Right-hand dynamometer53	-.64	.10	-.17	.73
14. Left-hand dynamometer54	-.68	.13	-.18	.80
16. Spirometer53	-.58	-.16	.23	.70
18. Height49	-.64	-.17	.19	.71
19. Weight55	-.67	-.17	.08	.79
20. Shop operations quality criterion ..	.55	.47	-.25	.01	.59
25. Son's mech. operations35	.14	-.23	-.35	.32
37. Interest Analysis Blank (new)50	.37	-.22	-.14	.46

TABLE 3
Transformation Matrix

	I	II	III	IV
I	.58	-.72	-.08	.39
II	.39	-.18	.34	-.83
III	.67	.59	-.44	.00
IV	.25	.32	.82	.39

TABLE 4
Rotated Factor Matrix

	I	II	III	IV	h^2
3. Packing Blocks05	-.11	.23	.65	.49
4. Card Sorting07	-.16	.11	.67	.49
5. Minn Spatial Relations02	-.01	.74	.26	.62
6. Paper Form Board	-.02	-.07	.70	.10	.51
7. Stenquist Picture I06	.39	.47	.23	.43
8. Stenquist Picture II	-.17	.28	.42	.37	.42
9. Minn. Assembly	-.03	.25	.71	.15	.58
11. Back dynamometer61	.48	.00	.18	.63
12. Right-hand dynamometer69	.49	-.06	-.07	.73
14. Left-hand dynamometer72	.52	-.10	-.05	.80
16. Spirometer83	.07	.05	-.10	.71
18. Height82	.05	.03	-.16	.70
19. Weight84	.20	.05	-.19	.79
20. Shop operations quality criterion ..	.00	.03	.75	-.19	.60
25. Son's mech. operations03	.32	.41	-.15	.29
37. Interest Analysis Blank (new) ...	-.02	.17	.66	.01	.47

numerical solution in the present study. As the factors are discussed variable by variable, the data will be presented in the form of factorial equations.¹

Factor I appears to be a size or a maturational factor. It is determined primarily by variables 16, 18, and 19.

Factor I—Size

	I	II	III	IV	U ²
11. Back dynamometer37	.23	.00	.03	.37
12. Right-hand dynamometer ..	.48	.24	.00	.00	.28
14. Left-hand dynamometer ..	.52	.27	(-) .01	.00	.20
16. Spirometer69	.00	.00	(-) .01	.30
18. Height67	.00	.00	(-) .03	.30
19. Weight71	.04	.00	(-) .04	.21

Approximately 70 per cent of the total variance of each of these variables is found in Factor I and no significant amount of variance is contributed by these variables to any other factor. Tests 11, 12, and 14, which suggest a strength factor, Factor II, actually have most of their common factor variance and approximately 50 per cent of their total variance in Factor I. This finding is of interest because in an analysis of data for 328 youths who were older than the present group it has been found that strength and size are independent of each other (2). Because of this, Factor I and Factor II are treated in the present study as independent of each other. The writer offers as additional justification for this treatment the consideration that no additional understanding of the organization of the variables would result from rigorously defining Factor II as highly correlated with Factor I.² Since the data of the present study do not call for a strength factor independent of the size factor, this independence can only be considered as hypothetical. It is reasonable to find size and strength highly correlated among young boys and to expect these variables to become increasingly independent as maturation is attained. It is hoped that the results of this study will have implications for the use of certain

¹ Factorial equations are more revealing than simple factor loadings because they not only show how much of the total variance of the test is due to each factor, but they also show how much of the variance is due to the common factors, i.e., how much (u^2) is unique to the test in the factorial equation. The u^2 values in the factorial equations are equal to the respect variance of the test.

² Actually the present data could be analyzed yielding 3 factors: size-strength, spatial ability, and a third factor. The present solution is somewhat "forced" and justified by the present common sense interpretation.

types of tests in the selection and guidance of young adults. It is hypothesized, therefore, that for such young adult groups body size and strength of the upper parts of the body are relatively independent of each other.

Factor II, the postulated strength factor, is of considerable interest in the present study because its variables do not contribute to the manual dexterity factor, Factor IV.

Factor II—Strength

	I	II	III	IV	U ^a
11. Back dynamometer37	.23	.00	.03	.37
12. Right-hand dynamometer ..	.48	.24	.00	.00	.28
14. Left-hand dynamometer52	.27	(-) .01	.00	.20

The contribution which Factor II makes to certain other variables such as the Son's Mechanical Operations variable and the Stenquist Assembly tests is meaningful insofar as strength of hands among boys would be expected to be associated with the use of the hands either as indicated directly by the Son's operations questionnaire or indirectly by the Stenquist Assembly tests which sample mechanical knowledge. The fact that variables 3 and 4, the manual-dexterity variables, do not contribute to this factor in any way is taken as additional evidence that manual dexterity is a classification of ability quite independent of other types of manual ability (6).

Factor III, the spatial relations factor is defined by the *Minnesota Mechanical Assembly Test*, the *Minnesota Paper Form Board Test* and the *Minnesota Spatial Relations Test*.

Factor III—Spatial Visualization

	I	II	III	IV	U ^a
5. Minn. Spatial Relations00	.00	.55	.07	.38
6. Paper Form Board00	.00	.49	.01	.50
7. Stenquist Picture I00	.15	.22	.05	.58
8. Stenquist Picture II	(-) .03	.08	.18	.14	.57
9. Minn. Assembly00	.06	.37	.02	.55
20. Shop operations quality criterion	.00	.00	.56	(-) .04	.40
25. Son's mechanical operations00	.10	.17	(-) .02	.71
37. Interest Analysis Blank (new) ..	.00	.03	.44	.00	.53

It is most interesting to observe that variable 20, the Shop Operations Criterion, has all of its common factor variance and over 50 per cent of its total variance in this particular factor. In addition, mechanical interests, 37, and Son's Operations in

mechanical activities, 25, are also highly correlated with this factor.

The importance of measures of spatial ability as indices of mechanical promise is strikingly indicated by the nature of this factor. Not only the criterion but interest in mechanical activities appears to be quite independent of the three additional factors which appear in this study and which might on an a priori basis be expected to contribute to a shop operations criterion of mechanical ability.

Factor IV is perhaps the most interesting factor in the present study.

Factor IV—Manual Dexterity

	I	II	III	IV	U ²
3. Packing Blocks00	(-) .01	.05	.42	52
4. Card Sorting00	(-) .03	.01	.45	51

It is defined by two tests which appear to call for a type of manual dexterity. However, these tests do not contribute to the spatial-visualizing factor which in the light of this study is the mechanical-ability factor. Perhaps more surprising is the fact that neither the strength nor the size factors contribute in any way to facility in manual dexterity as identified by this factor. Although manual dexterity is an ability which has long been considered as a definable attribute, prior to the investigations of this series its existence as a functional classification of ability had not been satisfactorily demonstrated. As a matter of fact the data presented in the present and in the preceding study may not be regarded as adequate to define satisfactorily an ability such as manual dexterity. This reservation is reasonable since block packing and card sorting were principal variables in defining this factor in both of these studies. Although the studies were done on two samples and the factor occurs in two different test batteries, its existence requires further demonstration.

Further Evidence for Identifying Manual Dexterity

In order to shed more light on the existence and the nature of this factor additional data were sought for further scrutiny. Data suitable for this purpose were found in the *Measurement*

of *Manual Dexterities* by Earle (1). He had published the intercorrelations of ten measures, nine of which were considered to be measures of manual dexterity. The tenth was a criterion of mechanical ability, the exact nature of which was not defined in his publication. The intercorrelations of these variables for a sample of 79 continuation Day-School boys are presented in Table 5³ and the nature of each of the performances is indicated below:

TABLE 5
*Intercorrelations for 79 Day-School Boys**

	I	II	III	VI ₁	VI ₂	VII	VIII	IX	XIII	Crit.
I										
II	.31									
III	.43	.29								
VI ₁	.14	.30	.24							
VI ₂	.13	.18	.20	.39						
VII	.16	.00	.24	.42	.25					
VIII	.30	.19	.31	.29	.25	.29				
IX	.17	.09	.21	.38	.33	.32	.39			
XIII	.04	.13	.02	.40	.29	.19	.26	.19		
Crit.	.20	.19	.11	.22	.12	.10	.24	.16	.08	

* Median age, 14 years and 5 months.

- Test I. Tapping movement of forefinger using wrist. A lever is tapped by the forefinger of the preferred hand while the non-preferred hand holds the apparatus.
- Test II. Tapping movements of several fingers in succession using wrist. The individual taps with each of the four fingers successively beginning with the little finger each time and tapping in order from the little finger to the index finger as quickly as possible.
- Test III. Twisting movements of finger and thumb with wrist action. In this test the individual is required to turn the barrel of the turn buckle until the eye is as far in the barrel as possible and then to reverse the direction of rotation and turn the barrel as quickly as possible until the eye is released.
- Test VI, part I: The individual is required to place 100 pegs into a 100-hole peg board, picking up one peg at a time.

³ Precise scoring methods used in securing these data are not specified by the author.

- part II: The individual is to fill three rows of the peg board, halting temporarily between rows.
- Test VII. The individual is requested to place the pegs in the peg board using the thumb and each finger (except the forefinger) in succession to pick up the pegs.
- Test VIII. Placing pegs in holes under manipulative difficulties.
- Part I: The individual takes all the pegs from one row of the peg board and then replaces them keeping them in his hand as he works;
- Part II: The individual extracts all the pegs from two rows of the peg board and then returns them keeping them in his hand during the process.
- Parts III and IV are conducted under such manipulative difficulties as maintaining the operating hand full of pegs while working.
- Test IX. Placing pegs in holes which are not visible. In this test a tactual exploration is made by the free hand, usually the left, while the right hand is used to pick up the pegs. The subject is not blindfolded but a screen is placed between him and the board. It is required that the subject feel for the hole each time with his left hand.
- Test XIII. Discrimination between fine and coarse textures by sense of touch. A screen is placed between the individual and a tactual board upon which strips of sand paper are placed and manipulated in such a fashion as to permit the individual to attempt to tactually recognize the match for several different grades of sand paper.

Tests IX and XIII are of particular interest because they are relevant to a question raised in the first paper in this series; it was found that the digit-symbol-substitution test, which would appear to call for no high degree of manipulative dexterity, was significantly correlated with the dexterity factor. The identity and the nature of the dexterity factor was put in

doubt by this finding. It appeared plausible that the dexterity factor is the old hand-eye coordination ability so commonly spoken of in earlier studies of mechanical testing. It was suggested that a hypothesis that manual dexterity calls for visual recognition and discrimination as well as a manipulative ability could be tested by screening the manipulative work from the testee's field of vision. Test IX suggests the possibility of a preliminary test of this hypothesis.

In order to examine this possibility and to seek further evidence of manual dexterity as a valid functional classification of ability, Earle's intercorrelations of the ten variables were submitted to a factor analysis. It was found that two centroid factors, Table 6, permitted a satisfactory reconstruction of the intercorrelation table. The two factors were then subjected to a single orthogonal rotation and the nature of the rotated factors, as shown in Table 6, appears to be meaningful.

TABLE 6

	<i>Centroid Factors</i>		<i>Rotated Factors</i>	
	I	II	A	B
I. Tapping47	-.44	.65	.06
II. Tapping41	-.30	.50	.11
III. Twisting51	-.32	.58	.17
VI. Pegs65	.31	.19	.70
VI. Pegs51	.24	.16	.54
VII. Pegs49	.27	.12	.55
VIII. Pegs (handicap) ..	.59	-.06	.44	.40
IX. Pegs (not visible) ..	.54	.18	.22	.53
XIII. Tactual41	.30	.05	.51
Criterion34	-.13	.33	.17

Factor A calls for the same class of operations as the ballistic or repetitive movement factor described in the first paper of this series. Earle's test III which calls for a simple, highly speeded twisting or twirling movement of the fingers is highly correlated with Factor A. This finding suggests that the simple repetitive movement of tapping involves the same ability as the more industrially significant repetitive twisting or twirling manual operations. Tests for the repetitive movement factor may conceivably be of considerable value in selecting workers for certain types of common industrial piece work employment.

Factor B seems to call for the same type of manual facility or manipulative ability which characterizes the dexterity factor revealed by the analyses of the Minnesota data. The tests differ from any of those used in the Minnesota study. One of them, test IX, does not permit the use of vision. Performance on test XIII depends upon accuracy of tactual discrimination. The analysis of Earle's data not only confirms the tendency for measures of manipulative ability to be positively intercorrelated but contributes to our understanding of this tendency. Tests calling for controlled placing, or adjusting movements, of the hands are interrelated as if they depended upon a single ability or capacity. This ability appears to be independent of the visual modality; it may be chiefly dependent upon the tactual and kinesthetic modalities.

The tendency for tests involving manual dexterity to be more highly correlated among themselves than with other tests is manifested in yet another context. Teagarden (5) has intercorrelated two Kent-Shakow scores, *Minnesota Spatial Relation*, two Minnesota *Rate of Manipulation* scores, and two scores from the *Cincinnati Pliers Test*. The tendency for the spatial visualization tests to form a correlation cluster different from the cluster formed by the dexterity tests is unmistakable.

Conclusions

Guidance experts and personnel technicians use tests which on the basis of the current statistical classifications are considered to be measures of steadiness, repetitive movement, dexterity, and spatial visualization. Yet, with the exception of the use of the spatial visualization tests, their procedures are justified chiefly on the basis of intuition and not on the basis of high correlations with external validating criteria, i.e., specific industrial performance. The literature abounds with evidence of the validity of spatial visualizing ability tests for prediction of mechanical work. Acceptable external validities of manual dexterity tests are more rare, however. In general, the most promising validity coefficients for manual dexterity tests have been obtained with ratings of supervisors as a criterion. Evidence of validity for the practical use of measures of repeti-

tive movement and steadiness are practically non-existent in the literature.

The suggestion is offered that the common failure to validate tests of factors other than spatial visualization and scholastic ability (which at the adult level may be fractionated into other abilities) is probably due to the nature of the criteria that have been employed. Most of the criteria that have been employed in the prediction of mechanical ability have been work samples prepared under unusual competition and other atypical conditions which appear to call for a much higher order of spatial visualizing judgment than manipulative ability, e.g., the criteria used in the Minnesota study. The so-called motor aspects of mechanical ability cannot be assumed to be of limited significance simply because their significance has not been rigorously demonstrated by suitable studies. If investigators employed such criteria as satisfaction in work, duration of employment in routine operations, speed of work, quality of specific operations, piece work output, breakage, fatigability and other factors of great practical significance in industrial operations, it might well be demonstrated that the motor abilities, particularly manipulative ability could, on the basis of demonstrated predicted value, be granted a significant rôle in guidance and selection procedures.

The term "mechanical ability" does not lend itself to adequate definition, however. In modern industrial employment there are innumerable different operations which involve the use of machines, tools and other mechanical contrivances. It appears likely that the successful prediction of satisfactory performance and good morale in these industrial activities is more dependent upon the development of adequate criteria than upon the invention of new ability tests. It is suggested that the greatest immediate progress in the field of mechanical ability testing depends upon extensive factor analytical studies of interrelationships of criteria of different phases of industrial operations and at different levels and types of work. Unfortunately such varied criteria would not be available for most groups of industrial workers; certain paid apprenticeship or training groups would probably be the most desirable subjects

for this research. It is only on the basis of such intensive research that mechanical ability may be satisfactorily defined. Definition of mechanical and manual work on any other basis is arbitrary and therefore not likely to be generally applicable.

The studies of the present series demonstrate the commonly observed tendency for intercorrelated psychological variables to form clusters, i.e., to permit a somewhat rigorous mathematical classification of the variables. The question which continually arises in a discussion of such studies as these is what significance may be ascribed to the classifications. The classifications which have been established by factor analysis could certainly be due to the sampling of the measures which are subjected to analysis. The sampling could be either a deliberately or an unconsciously obtained result.

The test samples could be a function of our culture. Perhaps the human organism is physically capable of an indefinite variety of response patterns. If this were so, the culture could be regarded as determining which response patterns are of practical significance and through this mode of influence the culture could also determine the pattern of performances sampled by current psychological tests. In addition to this selective effect, it is conceivable that a given culture may actually determine the development of abilities. Just as response patterns are elicited by life experiences, ability patterns may appear in a group of individuals in response to the exigencies of existence in the society. These cultural requirements may possess a structure which could be reflected in the organization of ability. The appearance of a pattern of ability among all of the individuals participating in a culture is expressed by the consistency in the society of intra-individual differences with respect to the various classes of ability. Currently the most satisfactory explanation of the development of intra-individual differences would rest upon learning theory. Learning theory is sufficiently well developed to enable us to envisage in a general way the manner in which differences in the environments of individuals could favor the development of intra-individual differences along lines which would reflect aspects of the culture.

The great diversity of culture has been observed from group

to group at different periods with respect to many important attributes. It has not been shown to the writer's knowledge, however, that the factorial pattern of ability varies meaningfully from culture to culture. Until this diversity has been demonstrated, the arresting possibility remains that the organization of ability may be a more or less standard pattern and a necessary consequence of the functional limitations of the human organism. A careful investigation of cultural differences in ability patterns appears to be necessary for the development of a science of human ability.

As previously mentioned, validity of tests for the various factors is directly dependent upon the type of criteria employed. Regardless of the nature of the criteria, however, it is frequently found that specially devised tests of rather anomalous factorial composition show higher validities than tests for known factors. The superior validity of tests which are specific to a task implies the existence of specific factors practically significant for the tasks. The ultimate significance of such hypothetical specific factors is unknown and probably rests in part upon the nature of their origin. Since the identifiable, stable factors are apparent among individuals at different age levels, it may be inferred that such abilities are relatively stable within the individual and therefore may be validly employed in long range predictions. It is possible, however, that specific abilities measured by certain tests are readily acquired in response to specific experiences and may have no great ultimate predictive value despite their value in predicting criteria established shortly after the application of the original test. More data establishing the long term predictive value of all of our tests are greatly needed. Long term predictions are always less reliable than immediate ones. This effect is doubtlessly due in part to the loss or acquisition of readily acquired, transient, specific abilities.

Summary

On the basis of factorial analyses of the Minnesota data and the examination of data of other studies of mechanical ability, it is apparent that a complete assay of an individual's potenti-

alities for all types of mechanical or manual work would call for measurement of at least the following attributes:

- | | |
|--------------------------|------------------------|
| 1. scholastic ability | 5. repetitive movement |
| 2. spatial visualization | 6. steadiness |
| 3. perceptual speed | 7. strength |
| 4. manual dexterity | 8. size |

The exact organization of the factors at different age levels requires further analytical study. The degree to which the importance of each of these abilities varies from job to job is unknown, but it is subject to critical determination.

REFERENCES

1. Earle, F. M. "The Measurement of Manual Dexterities." *Report of The National Institute of Industrial Psychology*, Vol. V. London: Aldwych House.
2. Hall, D. M. and Wittenborn, J. R. "Motor Fitness Test for Farm Boys." *Research Quarterly American Association Health and Physical Education*, XIII (1942), 432-443.
3. Harrell, W. "A Factor Analysis of Mechanical Ability Tests." *Psychometrika*, V (1940), 17-33
4. Patterson, D. G., Elliott, R. M., Anderson, L. D., Toops, H. G. and Heidebreder, E. *Minnesota Mechanical Ability Tests*. Minneapolis: The University of Minnesota Press, 1930.
5. Teagarden, L. "Manipulative Performance of Young Adult Applicants at a Public Employment Office, Part III" *Journal of Applied Psychology*, XXVI (1942), 754-769.
6. Wittenborn, J. R. "Mechanical Ability, Its Nature and Measurement, 1. An Analysis of the Variables Employed in the Preliminary Minnesota Experiment." *Journal of Educational and Psychological Measurement*, V (1945), 243-262.

SPEED AND LEVEL COMPONENTS IN TIME-LIMIT SCORES: A FACTOR ANALYSIS¹

WILLIAM M. DAVIDSON

1st Lt., Army Air Corps Classification Section, Biggs Field, Texas

AND

JOHN B. CARROLL

Lt. (jg), USNR, Bureau of Medicine and Surgery

WHETHER by force of tradition, or for reasons of expedience, it has been the practice to administer and score group tests of ability, aptitude, and achievement in such a way as to yield only a time-limit score, defined as the number of items correctly answered within a specified length of time. Thus, the time-limit score often becomes the sole measure of the behavior represented in a test. When a test is "validated" with respect to some external criterion, a time-limit score, rather than some other type of score, is most likely to be used as the measure which is correlated with the criterion. Likewise, in making a factor analysis of a battery of tests, one is most likely to use time-limit scores. It is the writers' belief that the indiscriminate use of time-limit scores is one of the more unfortunate characteristics of current psychological testing since the time-limit score of a test frequently represents *two* relatively independent aspects of behavior: (a) the amount the subject knows or can perform (or in certain cases, the level of difficulty which he can reach), and (b) the rate at which the subject works. Somewhat at variance with current usage, we shall identify these aspects of test behavior, respectively, by the terms *level* and *speed*.² By ignoring the possibility that these two aspects of test performance may play different rôles in any given situ-

¹ This paper is a revision of a thesis presented by the first-named author as a candidate for the M.A. degree at Indiana University 1943

² These terms were employed by Baxter (1), who was able to show a marked independence of speed and level in a single omnibus test of intelligence. The present study, in effect, extends Baxter's approach to tests of varying content.

ation, the applied psychologist runs the risk of obtaining validity coefficients lower than those which might be obtained if the level and speed components were correctly weighted in the prediction. For example, if the level score on a test has greater validity than the speed score in predicting a criterion, use of the time-limit score may tend to mask the potential validity of the test by introducing the "dead wood" variance of the speed component. In factor analysis there exists a real danger that the primary factors which are found in a set of time-limit scores may themselves be factorially complex, that is, that they may consist of both speed and level components. When these primary factors are correlated, as is frequently the case, one should not consider the hypothesis that the correlations indicate the presence of a general factor of intelligence until it is shown that they are *not* due to the presence of an underlying speed factor.

It is true that a logical distinction between speed and level elements in test performance has long been recognized. However, in practice it has been assumed that since these elements appear to be highly correlated they are merely different aspects of the same underlying entity and that consequently the distinction can be ignored. Furthermore, it has been assumed that in any case the normal exigencies of group test administration preclude any attempt to make separate measurements of speed and level. Without undertaking to review the literature on the problem, we believe that these assumptions will bear analysis.

The assumption that speed and level are different aspects of the same thing has arisen partly through confusion in terms and partly through misinterpretation of the experimental evidence. The most frequent error is that of identifying time-limit scores as "speed" scores and then proceeding to cite correlations between time-limit scores and scores obtained in unlimited time. The point has been missed that these correlations are spuriously high, since they rest on a part-whole relationship. The score obtained in unlimited time is equal to the time-limit score *plus* whatever the subject can accomplish in additional time. Moreover, the correlation between these scores is a function of the length of the time-limit, for obviously as the time-limit is

lengthened the time-limit scores become more similar to scores obtained in unlimited time. In any case, correlations as high as even .7 or .8 are still not high enough to rule out the possibility that a speed component, independent of level, exists in the time-limit scores.

The number of studies in which correlations have been obtained between rate-of-work scores and level scores is exceedingly small. The obtained correlations are seldom more than moderately high but even these have occasionally been cited as showing the fundamental identity of speed and level components in test performance.

These misinterpretations have usually occurred in connection with test performances in which the subjects vary considerably with respect to their ability to answer the items and in which the scores involve the number of items correctly answered. A particularly dangerous misinterpretation, however, is likely to arise in connection with tests in which the subjects vary *not* in their item-passing ability, but only in their rate of performance. One frequently cited study is that carried out by Paterson and Tinker (7), who came to the perfectly sound conclusion that when corrected for attenuation, the correlation between "work-limit" and "time-limit" scores on a speed-of-reading test is virtually perfect. The work-limit score was not the number of items correctly performed in unlimited time but, instead, the time taken to read all the paragraphs in the test. The time-limit score was the number of paragraphs read within a time-limit. What Paterson and Tinker showed, then, was that in the measurement of a *rate* of performance it makes little difference whether the scores are expressed in terms of performance-per-unit-of-time or time-per-unit-of-performance. A convenient paradigm is that of a runner's speed, which can be expressed either in terms of feet per second or in terms of seconds per foot. It is a mistake, however, to generalize the results of the Paterson and Tinker study by inferring that "work-limit" and "time-limit" scores in the usual sense will be highly correlated in situations where elements of test performance other than rate are measured.

With respect to the presumed impracticability of measuring

speed and level components separately within a time-limit, we can only point out that few attempts have been made to explore the problem. With ingenuity, it should be possible to devise relatively simple methods of making separate measurements even within a reasonable time-limit.

We would by no means assert that speed and level components of test performance are invariable entities from test to test. Rate of performance in one task may be completely independent of rate of performance in another task. Similarly, level components undoubtedly vary from test to test. The investigation reported here establishes the independence of several distinct types of speed, in addition to a general speed factor, and previous factorial investigations have isolated several types of level components (such as vocabulary knowledge, ability to solve problems expressed verbally, etc.).

We conclude this general introduction by making several recommendations in the fields of test construction and factor analysis. First, we suggest that persons responsible for the standardization and validation of tests experiment with the differential validities of speed and level scores and incorporate any significant findings in the directions for administering, scoring, and interpreting the tests. Investigations should be made of the possibility of restandardizing various published tests in terms of speed and level. Persons charged with selecting tests for use in given situations should give preference to tests which have been so standardized. Collateral experiments should meanwhile be directed towards discovering more efficient and reliable methods of measuring speed and level than, say, those employed in the present investigation.

Our second major recommendation is that in factorial studies aimed at discovering unitary abilities, tests should be represented by speed scores, level scores, or both, and that if time-limit scores are to be studied at all they should be treated in the manner exemplified in the investigation reported here.

The Experiment

In order to establish the linear independence of speed and level scores it was decided to study by factor analysis a matrix

of correlations between speed, level, and time-limit scores in a number of short mental ability tests. As in Baxter's study (1), speed scores were obtained as the number of seconds taken by the subject to work from the beginning to the end of the test, attempting every item once. Level scores were defined as the number of items correctly answered when the subject is allowed to take all the time he desires to try every item and to check over his work. Time-limit scores were defined as the number of items correctly answered within a prescribed time-limit.

The test battery consisted of the eight subtests of the *Revised Alpha Examination*, Form 5; the *Minnesota Speed of Reading Test for College Students*, Form A; and several tests which had been specially constructed for previous factorial investigations. These included Letter Grouping and Scattered X's, studied by Thurstone (8); and Phrase Completion and Disarranged Morphemes, constructed by Carroll (2, 3). The *Revised Alpha Examination* was used because its subtests appear to measure verbal, numerical, and reasoning factors, to judge from Guilford's analysis of the original Army Alpha test (4), and because it is somewhat more practicable to administer and score than the original Army Alpha. Letter Grouping and Disarranged Morphemes were included to aid in defining the domain of reasoning ability. Scattered X's was included to test the hypothesis that the Perceptual Speed factor (P) as measured by the test might be involved in some of the speed scores studied here. The *Minnesota Speed of Reading Test* was included because it was believed that speed of reading might be related to speed scores on mental tests which contain reading material.

Speed, level, and time-limit scores (as defined above) were obtained for each test or subtest in the battery, with three exceptions. For the *Minnesota Speed of Reading Test*, the only score obtained was the number of paragraphs marked, correctly or incorrectly, in the prescribed 6-minute time-limit. This score measures the speed aspect of performance on the test. The score on Scattered X's was the number of x's found and marked in 4 minutes; again, this score is primarily a measure of rate of performance. Phrase Completion had no time-

limit and was scored by means of a key the construction of which has been described in a previous article (3). Special instructions and procedures were devised to obtain speed, level, and time-limit scores on the same test.

A large clock with a sweep-second hand was placed in view of the subjects. First the subjects worked on a test for the prescribed time-limit, marking an "x" in the margin after the last answer written within the time-limit. They were then instructed as follows: "Continue working rapidly on the test to the end, but do not yet change any answers you have already written. As soon as you each individually finish the test, quickly look up at the clock and record at the bottom of the page the minutes and seconds you required to do the remainder of the test below the X. Do not stop too long on any one problem. You may guess at answers you don't know or leave blanks. Write the time before you go back to fill blanks or make corrections. After you record your time, you may take your red pencil and make any additions or corrections, but do not erase present answers." The students were allowed to work on each test until all had finished, except on the Disarranged Morphemes and Letter Grouping tests, where a few students were not able to finish within 23 and 18 minutes, respectively, after the time-limit. This procedure yielded a *time-limit* score, which resulted from the application of the prescribed scoring formula to all answers written in ordinary pencil up to the X marked by the subject. The *speed* score was the number of minutes and seconds recorded at the bottom of each test. The *level* score was the score on the entire test; if an answer in black was followed by a different one in red, the latter was taken as the answer for purposes of arriving at the level score.

The time-limits used for the subtests of the *Revised Alpha Examination* were those recommended by Tinker and Baker (9) for use with college students. For experimental purposes scores for two time-limits—2 and 4 minutes—were obtained for subtest 2, Arithmetical Reasoning.

The subjects were undergraduate students in elementary experimental psychology at Indiana University. The analysis of test scores was based upon 91 complete cases—12 men and 79 women.

The markedly skewed distributions of certain speed scores (on the subtests Addition, Common Sense, Same-Opposite, and Disarranged Sentences in the Revised Alpha) were made more

nearly normal by converting them to the reciprocals of the number of seconds. Scores on all variables were coded in ten or fewer class intervals. Hollerith procedures were used to obtain Pearsonian product-moment coefficients. No corrections for grouping or attenuation were applied to the coefficients. Before the correlation matrix was assembled for factor analysis, the level and time-limit scores on subtests Addition, Common Sense, and Disarranged Sentences were discarded, first, because most of the students made perfect scores in unlimited time, and second, because time-limit scores were very highly correlated with speed scores. Scattered X's was omitted from the analysis because it was little correlated with any other variable in the battery. It was therefore concluded that the Perceptual Speed factor as measured by Scattered X's is not significantly involved in the speed variables studied here.

The Factor Analysis

Level and time-limit scores, as defined here, are overlapping measures since the level score on a test can be regarded as equal to the time-limit score plus whatever additional correct answers the subject can give in time beyond the time-limit. There is likewise an obvious overlap between speed and time-limit scores since the faster the subject works the more items he has an opportunity to pass within the time-limit. It was believed that these factors of overlap would introduce spurious dimensions in the factor analysis if the correlation matrix were analyzed in the usual fashion. To put the matter differently, insertion of the time-limit scores would spuriously raise the communalities of the speed and level scores. A special method of factoring the matrix was suggested by Dr. L. R. Tucker. The *main* matrix, involving only speed and level variables, was analyzed in the usual way by the centroid method. All correlations between time-limit scores and speed scores or between time-limit scores and level scores were placed in a *subsidiary* matrix and factored separately. (See Table 1 for the correlations represented in the main and subsidiary matrices.) Correlations *among* time-limit variables were not analyzed at all. Essentially, the procedure involved locating the time-limit variables in the factor space

TABLE 1
The Correlation Matrices*

		"Main" matrix															"Subsidiary" matrix											
		5	6	7	8	9	10	11	12	13	14	16	18	20	21	22	23	24	25	35	27	28	30	32	33	34	36	38
5																												
6																												
7																												
8																												
9																												
10																												
11																												
12																												
13																												
14																												
16																												
18																												
20																												
21																												
22																												
23																												
24																												
25																												
35																												

* This table presents only the correlations used in the factor analysis. The names of the variables are given in Table 4. All entries have been multiplied by 100 to eliminate the decimal point.

defined by the speed and level variables. Factor loadings for the variables in the subsidiary matrix were obtained by summing the columns of correlations or residuals; the product of the column sum and the value $(\Sigma r)^{-1}$ used to compute the m th factor loadings for the main matrix was the m th factor loading for the subsidiary matrix variable. Residuals in the subsidiary matrix were computed and treated in the usual way.

TABLE 2
The Centroid Matrix

Test	I	II	III	IV	V	VI	h^2
5	.32	-.37	-.10	-.18	-.05	.27	.36
6	.43	-.40	-.09	.35	.18	-.18	.54
7	.71	.05	-.41	-.18	-.17	-.11	.75
8	.62	.10	-.34	-.16	-.04	.23	.59
9	.65	.12	-.28	-.16	-.04	-.05	.52
10	.63	-.36	-.10	-.18	.15	.10	.60
11	.71	-.06	-.21	-.14	.07	-.07	.58
12	.76	-.21	-.05	-.13	.05	-.06	.65
13	.64	.18	-.11	.32	-.32	-.02	.66
14	.39	-.16	-.23	.32	-.04	-.13	.35
16	.56	-.24	.29	.19	-.13	-.10	.52
18	.48	.44	.03	-.05	.32	-.04	.53
20	.67	-.12	.42	-.22	-.18	-.05	.72
21	.40	.18	.22	-.20	.15	-.24	.36
22	.61	.11	.33	-.17	-.22	-.14	.59
23	.65	.34	.28	.16	-.18	-.13	.70
24	.26	.06	.24	.21	.21	.31	.31
25	.51	.14	.35	.05	.10	.12	.43
35	.58	.17	-.34	.05	.22	-.17	.56
27	.69	-.23	.26	.22	-.05	-.22	.70
28	.58	-.25	.36	.07	.02	-.06	.55
30	.62	.44	-.23	-.20	.13	.26	.77
32	.70	-.28	.23	-.25	-.09	-.01	.69
33	.71	-.13	-.20	-.18	.15	-.17	.64
34	.75	.15	-.11	.03	-.01	.02	.60
36	.72	.26	.15	.26	-.07	-.09	.69
38	.62	.13	.25	.18	.17	.10	.54

As shown in Table 2, six centroid factors were extracted. The centroid matrix for the *main* correlational matrix was the basis for the rotation to simple structure, which was accomplished by Tucker's semi-analytical method (10) in five trials. The transformation matrix (Table 3) was used to obtain the final rotated matrix (Table 4) both for the *main* and the *subsidiary* matrix variables. The time-limit scores did not in any way influence the rotational procedures; nevertheless, the vec-

TABLE 3
The Transformation Matrix

	A	B	C	D	E	F
I	.33	.30	.16	.09	.18	.42
II	-.77	.09	-.16	.11	.32	-.11
III	.29	.73	-.24	-.18	.12	-.84
IV	.04	.29	.89	-.03	.20	-.05
V	.38	-.54	.07	.74	.43	-.15
VI	.27	.00	-.30	-.63	.79	.28

tors for these scores were found to fit well in the simple structure already established by the speed and level scores. Table 5 shows the correlations between the primary factors.

TABLE 4
The Rotated Factorial Matrix

Test Variable	A	B	C	D	E	F
<i>Speed Scores:</i>						
5 Addition38	-.04	-.13	-.18	.06	.34
6 Arith. Reasoning46	.02	.54	.28	-.06	.22
7 Common Sense	-.02	-.06	.04	.10	-.09	.63
8 Same-Opposite06	-.08	-.05	-.06	.25	.62
9 Disarr. Sentences04	.04	.02	.10	.04	.43
10 Number Series53	-.05	-.02	.09	.09	.41
11 Verbal Analogies21	-.01	.06	.16	.04	.46
12 Directions38	.10	.05	.13	.02	.36
13 Disarr. Morphemes	-.07	.39	.34	-.13	.10	.36
14 Letter Grouping13	.04	.46	.12	-.08	.34
<i>Level Scores:</i>						
16 Arith. Reasoning ..	.38	.50	.26	-.08	-.05	-.01
18 Same-Opposite	-.07	.05	-.01	.34	.32	.06
20 Number Series31	.53	-.16	-.14	-.03	-.02
21 Verbal Analogies ..	.02	.14	-.11	.30	-.02	-.10
22 Directions10	.51	-.10	-.04	-.05	-.06
23 Disarr. Morphemes ...	-.08	.61	.18	-.02	.10	-.04
24 Phrase Completion28	.21	.08	-.06	.46	-.06
25 Letter Grouping ..	.22	.38	.00	.01	.31	-.09
35 Speed of Reading02	-.16	.26	.38	.09	.46
<i>Time-Limit Scores:</i>						
27 Arith. Reasoning .. (2')	.43	.46	.32	.08	-.06	.06
28 Arith. Reasoning .. (4')	.50	.46	.10	-.02	.04	-.08
30 Same-Opposite ... (1')	-.08	-.13	-.16	.14	.46	.46
32 Number Series .. (2½')	.48	.34	-.13	-.08	-.04	.14
33 Verbal Analogies .. (2')	.30	-.09	.05	.30	-.04	.42
34 Directions10	.18	.16	.10	.18	.38
36 Disarr. Morphemes (8')	.04	.44	.28	.08	.18	.11
38 Letter Grouping .. (8')	.24	.36	.17	.07	.37	.02

TABLE 5
Correlations between the Primary Factors

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
<i>A</i>	1.000	-.044	-.082	-.035	-.242	.002
<i>B</i>	-.044	1.000	-.422	.688	.052	.635
<i>C</i>	-.082	-.422	1.000	-.460	.051	-.401
<i>D</i>	-.035	.688	-.460	1.000	.130	.516
<i>E</i>	-.242	.052	.051	.131	1.000	-.036
<i>F</i>	.002	.635	-.401	.516	-.036	1.000

Interpretations of the Factors

In interpreting the factors we follow the arbitrary rule that a projection larger than .30 indicates a significant loading of a test on a factor.

The variables having projections of .30 or greater on factor *A* are ranked below in order of size of projection. Significant projections on other factors are also given.

No.	Test Variable	Projections	
		<i>A</i>	Other factors
10	Number Series (speed)53	.41 <i>F</i>
28	Arithmetical Reasoning (4' time-limit) .	.50	.45 <i>B</i>
32	Number Series (time-limit)48	.34 <i>B</i>
6	Arithmetical Reasoning (speed)46	.54 <i>C</i>
27	Arithmetical Reasoning (2' time-limit)	.43	.45 <i>B</i> ; .32 <i>C</i>
5	Addition (speed)38	.34 <i>F</i>
12	Directions (speed)38	.35 <i>F</i>
16	Arithmetical Reasoning (level)38	.50 <i>B</i>
20	Number Series (level)31	.53 <i>B</i>
33	Verbal Analogies (time-limit)30	.30 <i>D</i> ; .42 <i>F</i>

Most of these tests obviously have to do with simple arithmetical computation; factor *A*, then, appears to be the number factor *N* identified in previous studies. The speed scores of Arithmetical Reasoning and Number Series have higher loadings than the corresponding level scores. The interpretation may be offered that for college adult subjects this factor refers to the *speed* aspect of computational behavior. The level of competence of these subjects is such that accuracy in arithmetic plays only an incidental rôle, although rapid arithmetical ability appears to facilitate the correct performance of the relatively complicated tasks set in Arithmetical Reasoning and Number Series. The presence of Directions (speed) on this factor becomes understandable when it is noted that a con-

siderable share of the items involve numbers and numerical operations.

Factor *B* has the following tests:

	<i>B</i>	Other
23 Disarranged Morphemes (level)61	—
20 Number Series (level)53	.31 <i>A</i>
22 Directions (level)51	—
16 Arithmetical Reasoning (level)50	.38 <i>A</i>
27 Arithmetical Reasoning (2' time-limit) ..	.46	.43 <i>A</i> ; .32 <i>C</i>
28 Arithmetical Reasoning (4' time-limit) ..	.46	.50 <i>A</i> ; 10 <i>C</i>
36 Disarranged Morphemes (time-limit)44	—
13 Disarranged Morphemes (speed)39	.34 <i>C</i> ; 36 <i>F</i>
25 Letter Grouping (level)38	.31 <i>E</i>
38 Letter Grouping (time-limit) ..	.36	.37 <i>E</i>
32 Number Series (time-limit)34	.48 <i>A</i>

In previous factorial studies tests similar to those represented above have been identified as tests of reasoning ability. Of the eleven variables listed, ten are directly or indirectly measures of *level* of ability (time-limit scores being regarded as a function of both level and speed). In the light of these considerations, factor *B* may be identified as a Level of Reasoning factor. The present battery is too limited to indicate the relation of this factor to the *inductive* and *deductive* reasoning factors which have been indicated in previous studies. The presence of the speed score of Disarranged Morphemes on this factor is interesting. In contrast to other tests in this battery, Disarranged Morphemes is of such a nature that it is almost impossible for a subject to be satisfied with an incorrect answer, the subject either solves an item correctly or is forced to skip it. Consequently, speed of performance in this task would be almost perfectly related to ability to answer the items if the subjects did not differ in their willingness to skip items. Because of this inherent connection between speed and level aspects of performance on the Disarranged Morphemes test, it is not surprising to find the speed score present on the Level of Reasoning factor. Parenthetically, we may say that there are several subtle problems in this area which this study has not been designed to handle. For example, one would like to know how the speed-level relationship varies with the difficulty of the task and whether the relationship in the case of multiple-choice tests is essentially different from that in the case of tests where

the subject is forced by the nature of the test to answer correctly or not at all.

Factor *C* is represented by the following test variables:

	<i>C</i>	Other
6 Arithmetical Reasoning (speed)54	.46 <i>A</i>
14 Letter Grouping (speed)46	.34 <i>F</i>
13 Disarranged Morphemes (speed)34	.39 <i>B</i> ; .36 <i>F</i>
27 Arithmetical Reasoning (2' time-limit) ..	.32	.43 <i>A</i> ; .46 <i>B</i>

The tests represented here are reasoning tests also found in factor *B*. Factor *C*, however, is constituted by measures of *speed*. *Level* is not independently represented at all. Factor *C* may hence be regarded as a Speed of Reasoning factor. As will be shown later by multiple regression techniques, the time-limit scores of these reasoning tests are much more heavily weighted with level than with speed. It is not surprising that only one time-limit score (from the 2' time-limit on Arithmetical Reasoning) appears on factor *C*. The 4' time-limit score on this test has a loading of only .10 on *C*.

It is of interest to note from Table 5 that there is an appreciable negative correlation between factors *B* and *C*. This probably indicates that when other factors are ruled out, those who are hasty in performing these reasoning tests are likely to be inaccurate.

Factors *D* and *E* lack definition in the present limited battery. They are represented by the following variables:

Factor <i>D</i> :	<i>D</i>	Other
35 Speed of Reading38	.46 <i>F</i>
18 Same-Opposite (level)34	.32 <i>E</i>
21 Verbal Analogies (level)30	
33 Verbal Analogies (time-limit)30	.30 <i>A</i> , .42 <i>F</i>

Factor <i>E</i> :	<i>E</i>	Other
24 Phrase Completion46	
30 Same-Opposite (time-limit)46	.46 <i>F</i>
38 Letter Grouping (time-limit)37	.36 <i>B</i>
18 Same-Opposite (level)32	.34 <i>D</i>
25 Letter Grouping (level)31	.38 <i>B</i>

Factor *D* may perhaps be characterized as a verbal reasoning factor which emphasizes formal relationships such as those of antonymity, genus-species, etc. Factor *D* is highly correlated with factor *B*, the Level of Reasoning factor. Were it not for the presence of both the level and time-limit scores of Letter

Grouping on the factor, factor *E* might readily be interpreted as the verbal factor identified in previous studies.

Factor *F* is represented by the following variables:

	<i>F</i>	Other
7 Common Sense (speed)63	—
8 Same-Opposite (speed)62	—
11 Verbal Analogies (speed)46	—
35 Speed of Reading46	.38 <i>D</i>
30 Same-Opposite (time-limit)46	.46 <i>E</i>
9 Disarranged Sentences (speed)43	—
33 Verbal Analogies (time-limit)42	.30 <i>A</i> ; .30 <i>D</i>
10 Number Series (speed)41	.53 <i>A</i>
34 Directions (time-limit)38	—
12 Directions (speed)36	.38 <i>A</i>
13 Disarranged Morphemes (speed)36	.39 <i>B</i> ; .34 <i>C</i>
5 Addition (speed)34	.38 <i>A</i>
14 Letter Grouping (speed)34	.46 <i>C</i>

Every one of these variables involves either a direct or an indirect measure of speed. It is also true that with one exception every speed score in the battery appears in the above list. Only four time-limit scores are absent: those of Arithmetical Reasoning, Number Series, Disarranged Morphemes, and Letter Grouping, and in these tests it can be shown that speed contributes little to the time-limit scores. Hence it may be concluded that factor *F* is a general speed factor involving rate of work in performance of tasks of the sort found in intelligence tests. The factor is similar to a general speed factor found in some of Holzinger's studies (5, 6). The content of a test does not seem to play any rôle in determining the loading of its speed score on factor *F*, since tests of verbal, numerical, and reasoning abilities all appear in the above list. No definite conclusions can be drawn from the present data, however, as to whether this factor extends to both easy and difficult tasks.

The presence of Speed of Reading on factor *F* might lead one to suspect that speed of reading is fundamentally involved in this factor. However, some of the tests whose speed scores measure the factor (e.g., Addition, Number Series, and Same-Opposites) do not have items containing connected text-material where a speed of reading factor could be expected to operate. It appears, therefore, that an individual's reading speed is partly a function of some more general speed factor.

Multiple-Correlation Analysis

Although it is believed that factorial techniques provide a more complete and concise analysis of the data, multiple-correlation techniques can be used to evaluate the independent rôles of speed and level in determining the variance of time-limit scores. Table 6 presents such an analysis for all tests in the

TABLE 6

Beta-Coefficients and Multiple Correlations in the Prediction of Time-Limit Scores (T) from Speed (S) and Level (L) Scores

Test	Zero-order Correlations			Relative Contributions of		$R_{T \cdot SL}$
	r_{TS}	r_{TL}	r_{SL}	Speed $\beta_{TS \cdot L}$	Level $\beta_{TL \cdot S}$	
Alpha Examination:						
1. Addition809	.344	.220	.770	.174	.826
2. Arith. Reas. (2')482	.797	.439	.165	.725	.811
2. " " (4')323	.712	.439	.012	.706	.711
3. Common Sense830	.296	.187	.804	.146	.843
4. Same-Opposites734	.616	.327	.596	.421	.835
5. Disarranged Sentences790	.518	.308	.698	.303	.842
6. Number Series673	.766	.485	.394	.575	.840
7. Verbal Analogies831	.392	.338	.788	.125	.840
8. Directions649	.566	.415	.500	.358	.724
Disarranged Morphemes596	.782	.564	.227	.653	.804
Letter Grouping451	.726	.137	.358	.677	.808

battery for which speed, level, and time-limit scores were obtained. The beta-coefficients indicate the relative contributions made by speed and level in predicting time-limit scores. In some tests, such as Arithmetical Reasoning, the time-limit scores are chiefly a function of the level of item difficulty that can be mastered by the subject, while in other tests, such as Common Sense, the time-limit scores are primarily measures of the subject's rate of work. In still other tests, such as Same-Opposite, speed and level are about equally weighted in the time-limit score. These relationships depend to some extent on the particular time-limits which had been set.

Even where the correlation between time-limit and level scores is fairly high the contribution of an independent speed component to the time-limit score is sometimes fairly large (e.g., in the case of Letter Grouping). The multiple correla-

tions in Table 6 are the correlations obtained in the prediction of time-limit scores from level and speed scores. These multiple correlations are in some cases considerably higher than the corresponding zero-order coefficients. Nevertheless, there remains in each case a certain amount of specific (unpredicted) variance in the time-limit score which would militate against the prediction of level scores, for example, from a weighted combination of speed and time-limit scores.

Summary

A number of relatively simple group mental tests were administered to 91 college students in such a way as to yield three types of score: *speed*, *level*, and *time-limit*. Speed scores represented the time required to attempt every item once; level scores represented the number of items correctly answered in unlimited time; and time-limit scores were the number of items correctly answered in a prescribed time-limit. Factor analysis revealed that in all cases speed scores were linearly independent of level scores and that time-limit scores could be represented as factorially complex measures having loadings on both speed and level dimensions of ability. Of the factors which were identified several were similar to verbal, numerical, and reasoning factors isolated in previous factorial studies. In the domain of reasoning ability both level and speed factors were identified. A general speed factor involving nearly all of the speed scores was found. It is concluded that because of their factorial complexity, time-limit scores should be used with considerable caution both in factorial studies and in studies involving the prediction of criteria.

REFERENCES

1. Baxter, B. J. "An Experimental Analysis of the Contributions of Speed and Level in an Intelligence Test." *Journal of Educational Psychology*, XXXII (1941), 285-296.
2. Carroll, J. B. "Knowledge of English Roots and Affixes as Related to Vocabulary and Latin Study." *Journal of Educational Research*, XXXIV (1940), 102-111.
3. Carroll, J. B. "A Factor Analysis of Verbal Abilities." *Psychometrika*, VI (1941), 279-307.
4. Guilford, J. P. *Psychometric Methods*. New York: McGraw-Hill Book Co., 1936. Chapter XIV.

5. Holzinger, K. J. *Preliminary Report on Spearman-Holzinger Unitary Trait Study*. No. 4. Chicago: Statistical Laboratory, Department of Education, University of Chicago, 1935.
 6. Holzinger, K. J. and Swineford, F. "A Study in Factor Analysis: the Stability of a Bi-Factor Solution." *Supplementary Educational Monographs*, No. 48, 1938
 7. Paterson, D. G. and Tinker, M. A. "Time-Limit Versus Work-Limit Methods." *American Journal of Psychology*, XLII (1930), 101-104
 8. Thurstone, L. L. "Experimental Study of Simple Structure." *Psychometrika*, V (1940), 153-168.
 9. Tinker, M. A. and Baker, K. H. *Introduction to Methods in Experimental Psychology*. New York: D Appleton-Century Co., 1938.
 - 10 Tucker, L. R. "A Semi-Analytical Method of Factorial Rotation to Simple Structure." *Psychometrika*, IX (1944), 43-68.
-

THE USE OF AN OBJECTIVE TEST IN PREDICTING RHETORIC GRADES

IRWIN A. BERG, GRAHAM JOHNSON, AND ROBERT P. LARSEN

University of Illinois

WHILE a passing grade in rhetoric or in an equivalent course is required of freshmen in virtually all colleges and universities, many institutions are exempting from the course those students who pass a proficiency examination. In addition, such proficiency examinations are sometimes used to determine whether a student should be admitted to the regular course in rhetoric or assigned to a special non-credit rhetoric class.

At the University of Illinois all entering freshmen are required to take a rhetoric placement examination. Those students whose performance is high are granted credit in Rhetoric 1 without taking the course. Those whose test performance is low are required to take Rhetoric 0, a non-credit course. Provision is made for early transfer to Rhetoric 1 of any Rhetoric 0 students whose classroom work proves to be at the level found in Rhetoric 1. All other students are entered in Rhetoric 1, the usual college course. In addition, a recent action of the University's Board of Trustees makes reasonable proficiency in written English a graduation requirement. Students earning grades of "C" or "D" in Rhetoric 2 are required to pass a special examination or to pass a third course in Rhetoric before being granted a bachelor's degree.

The *Rhetoric Placement Examinations* at the time of this study in October, 1943, consisted of an objective test¹ and an impromptu theme written in the examination room. The actual decision as to whether students were assigned to Rhetoric 0 or Rhetoric 1 or passed for proficiency was made on the basis of the quality of the impromptu theme. Two rhetoric

¹ Cooperative Test A: *Mechanics of Expression*, Form Q.

staff members individually evaluated each theme for allocation to one of the three groups. If the two instructors disagreed, the theme was given to a third instructor who made the final decision. The third instructor could consult the objective test score in such contested cases when the objective test results were available. The objective tests were scored by an International Business Machines electrical scoring machine.

The object of this study is to make inquiry into the usefulness of an objective test in predicting rhetoric achievement in college. Upon initial examination it may appear that such an inquiry could not prove fruitful since the rhetoric course grade is mainly determined by grades on what is often thought of as subjectively evaluated compositions. Thus a sample composition graded in the same manner might be considered an eminently more suitable predictive instrument. As Kelley and Roberts put it:

"We have found that ability to detect and correct errors in exercises is not always accompanied by the ability to avoid similar or worse errors in original composition, and that, conversely, students really proficient in composition may have indifferent success on a problem-solving type of test. We hold firmly the conviction that *a student's degree of proficiency in writing can be determined only by a demonstration of that proficiency, in writing.*"²

Yet the advantages of rapid scoring which could be done by persons who are not necessarily rhetoric instructors, together with the advantages of objectivity of score, would make the use of a suitable objective test an extremely practical measuring tool.

Two groups were used in this study. Group 1 numbered 372 students and group 2 numbered 166. Both groups were composed of freshmen entering the College of Liberal Arts and Sciences in the fall of 1943. The groups were not at first combined because each was tested in a different room and by different examiners, although the day and hour were the same for both groups. As will be noted later this precaution was unnecessary; consequently only in Table 2 are the groups sepa-

² Kelley, Cornelia and Roberts, Charles "Rhetoric Proficiency Tests" *Illinois English Bulletin*, XXXI (1944), 2.

rated. In analyzing the data from these groups Pearson product-moment correlations and other simple statistical measures were used. The grades of Rhetoric 0 students were lowered one grade point (i.e., from B to C) and the grades of students who were passed for proficiency in rhetoric were tabulated as B. This was in accordance with the recommendation made by members of the rhetoric staff.

Results

TABLE 1

Mechanics of Expression Test Scores in Relation to Grades in Rhetoric

Rhet. 1 Grade	N	Mean	S.D.	C.R.
A	25	142.7	9.8	10.5 (A & B)
B	154	117.6	17.4	
C	216	98.4	20.2	4.6 (C & D)
D	43	83.1	19.9	
E	8	73.3	29.5	.9 (D & E)
All Grades Rhet 1 .	446	105.4	24.2	3.2 (D & Rhet 0)
Rhetoric 0 . .	29	69.4	16.3	
Pass Proficiency . .	62	134.5	15.3	3.0 (A & Prof.)

TABLE 2

Correlation of Grades with Test Scores

	Group 1		Group 2	
	N	r	N	r
Mech. Exp. Scores and all Rhetoric grades* . .	372	.683	166	.696
Mech. Exp. Scores and Rhet. 1 grades only† . .	350	.639	159	.634
Mech. Exp. Scores and Rhet. 1 grades without proficiency group‡	313	.618	133	.607
Mech. Exp. Scores and Grade Point Average for all courses§	342	.541	162	.532
Mech. Exp. Scores and A.C.E. Total Scores . .	342	.659	162	.657
A.C.E. Total Scores and Rhetoric 1 grades¶ . .	302	.442	139	.465

* Grades of students "passed for proficiency" in Rhetoric 1 are taken as "B" Rhetoric 0 grades are lowered one grade point Mech. Exp. is the abbreviation for Cooperative Test A: *Mechanics of Expression*, Form Q

† Rhetoric 0 grades are not included

‡ Rhetoric 0 and "passed for proficiency" grades not included

§ Only those students who earned grades in courses totaling 12 or more semester hours are included.

|| A.C.E. refers to the *American Council on Education Psychological Examination*, 1940 edition

¶ Rhetoric 0 grades are not included Some Rhetoric 1 students did not take the A.C.E. examination

TABLE 3
*Types of Errors Checked in Early Freshman Compositions of 147
 Rhetoric 1 Students**

Type of Error	Total No. of Violations	Percentage of Students Making One or More Errors
1. Grammar (i.e., sentence fragment, incorrect tense or mood)	364	33.9
2. Mechanics (i.e., capitalization, italics)	298	44.1
3. Punctuation (i.e., superfluous comma, hyphen in compounds)	1153	50.2
4. Spelling	703	86.0
5. Diction (i.e., exactness, wordiness, faulty idioms)	1184	75.9
6. Coherence (i.e., word order, dangling modifiers, parallelism)	636	39.7

* Summarized from Johnson, W. G. and Mathews, E. G. Errors most frequently checked in early freshman compositions. *Illinois English Bulletin*, XXXI (1944), 1-8

Discussion

The curious aspect of these data is not that the objective test used in this study is a good predictor of achievement in Rhetoric 1. Admittedly, correlations above .60 between course grades and pre-test scores are not common. The more pertinent question is *why is the correlation so high?* Rhetoric 1 is a course in which the final grade is largely determined by grades earned on written themes in which the instructor's evaluation includes what may be considered subjective aspects such as *triteness*, *wordiness*, or *lack of logic*. The purely objective test, on the other hand, measures only basic skills in grammar, spelling, punctuation, and capitalization. There can be little doubt that the objective test does bear a clear relation to grades earned. In Table 1 it will be seen that there is a progressive decrease in the mean objective test scores by grades from "A" through "E." Rhetoric 0, composed of students deemed too poorly prepared to enter Rhetoric 1, has the lowest mean of all categories. The group "passed for proficiency" has a mean score between those of the "A" and "B" grade students. Several Rhetoric 1 instructors agreed beforehand that the proficiency students would probably fall at this level. The critical ratios of the differences between the test means of most grade categories indicate that the differences are highly significant.

An inference may be drawn from the data in Table 3 which partially explains why the predictive value of an objective test should be high for a course the content of which is ordinarily thought of as being subjectively graded. It will be noted that the majority of the errors of early freshman compositions were largely objectively ascertained errors, i.e., errors of spelling, punctuation, mechanics, and grammar. Significantly, it is this type of error which the objective test measures. In practice, then, the majority of the errors checked are detected objectively.

The weight, or value, rhetoric instructors assign to purely objective errors is probably quite high. A hypothetical case may demonstrate the importance of the value assigned to a given error. Suppose that two students each have exactly the same number of errors checked on their themes. The first student makes errors only of exactness and wordiness while the errors of the second student are errors in spelling, such as *alright*, and a series of errors such as *between him and I*, *he come around the corner*, etc. It is likely that the rhetoric instructor would consider the latter errors as abominations not to be lightly dismissed. Presumably the grade of the first student would be significantly higher than that of the second although the number of errors was the same in either case.

In other instances the instructor may grade compositions on a purely objective basis because he can do little else. Most of the themes may, at times, approximate each other in errors of diction, coherence, and organization while the variation between compositions in spelling, punctuation, etc., is marked. Also, the instructor may be influenced by the fact that such errors are more easily detected and are less likely to be contested by the student.

Another interpretation of these data might propose that there exists a parallel development of the mastery of English mechanics and of effectiveness in expression. Thus, skill in mechanics would tend strongly to be associated with variety of sentence structure, freshness of treatment, and superior organization. Conversely, lack of mechanical accuracy would tend strongly to be associated with a less effective presentation

of material. It might then be said that rhetoric instructors are accurate in evaluating both the mechanics and the effectiveness of composition. Since both aspects of writing are positively correlated, a false impression is given that the purely objective errors are emphasized. This interpretation would be strengthened by the fact that early freshman compositions, such as those recorded in Table 3, are very carefully scored for mechanical errors. Because most students would presumably improve in mechanics, the rhetoric instructor could place greater emphasis, when scoring later compositions, upon interest, originality, and freshness of treatment. This explanation would probably be favored by rhetoric staff members.

But it must be emphasized that, while such parallel development may exist to some extent, instructors would agree closely on scoring for mechanical accuracy but not on scoring for originality or superiority in organization. Yet the correlation between a test of basic mechanical skills given at the beginning of the semester with final rhetoric course grades at the end of the semester is almost .70. It would seem that if the more or less subjectively scored aspects of compositions are given much weight the correlation should be considerably lower, since variability in scoring is greater.

In general, it may be said that the high-school preparation of many students is inadequate insofar as mastery of basic skills in the mechanics of expression is concerned. It may be that a universal college admission requirement of four years of high-school English would improve the situation. Perhaps a required level of proficiency on a rhetoric test could be made part of the admission procedure. Such measures could be adopted only if it were agreed that the main function of a college rhetoric course is practice in English composition, not drill in punctuating such compositions. It is probable that the predictive value of an objective test in rhetoric will remain high as long as many students enter college with inadequate grounding in English mechanics. The correlation of such tests with grades should become progressively lower as students enter rhetoric classes with more thorough preparation in English.

This lack of thorough training in so-called "drill subjects"

may be reflected in other areas. It is not uncommon to encounter students, for example, who do not know the multiplication table. Occasionally one may discover a college student who has not learned the exact order of the letters of the alphabet. Whether or not mastery of "drill subjects" should be expected of entering college students cannot be discussed here. But it seems clear, in the case of mechanics of English at least, that such lack of proficiency does exist.

Summary

Data from 538 University of Illinois Liberal Arts College freshmen in Rhetoric 1 were analyzed. It was found that the scores of an objective test in the mechanics of expression correlated as high as .69 with final grades in rhetoric. Critical ratios of the differences in the mean objective test scores of students earning final grades of "A," "B," "C," etc., were significant. The question was raised as to why course grades which were presumably determined subjectively should correlate highly with scores on an objective test. The explanation advanced was that the preparation of many students was such that the rhetoric instructor was forced to grade largely on the basis of errors in mechanics. It was further suggested that instructors probably view more sternly objectively ascertained errors such as *he don't* and *alright* than the more subjectively determined errors such as triteness and freshness of treatment. Also, it was considered possible that the instructor may have been influenced when grading English compositions by the fact that purely objective errors are more easily detected and that the student is less likely to contest a grade based upon objective errors. The possibility that a parallel development exists between mastery of English mechanics and effectiveness of expression was considered but judged to be inadequate as an explanation.

A QUICK GRAPHIC METHOD FOR PRODUCT MOMENT ' r '

WILLIAM LEROY JENKINS

Lehigh University¹

THE product-moment coefficient of correlation can be determined graphically in a fraction of the time required to compute it mathematically. With large samples the two methods give virtually identical results. With small samples graphically determined coefficients appear to be about as representative of the relationship in the total population from which the sample was drawn.

The graphic method can be applied directly to raw data without grouping into class intervals. The correctness of the solution can be readily checked by inspection.

The method depends on the following relation between the coefficient of correlation (r) and a ratio (J) which can be determined graphically:

$$J = \sqrt{\frac{1+r}{1-r}}$$

Procedure (See Figure 1)

0. Plot the scattergraph directly from the raw data without grouping.

1. Move a straightedge from the top of the scattergraph, keeping it parallel to the x -axis until 16% of the plotted points show above the straightedge. Through the latest point to appear draw a line parallel to the x -axis.

2. Move a straightedge from the right side of the scattergraph, keeping it parallel to the y -axis, until 16% of the plotted

¹ On leave until January 1, 1946, with Columbia University, Division of War Research, Submarine Training Section, Box 34, Submarine Base, New London, Connecticut.

points show to the right of the straightedge. Through the latest point to appear draw a line parallel to the y -axis.

3. Move a straightedge from the bottom of the scattergraph, keeping it parallel to the x -axis, until 16% of the plotted points show below the straightedge. Through the latest point to appear draw a line parallel to the x -axis.

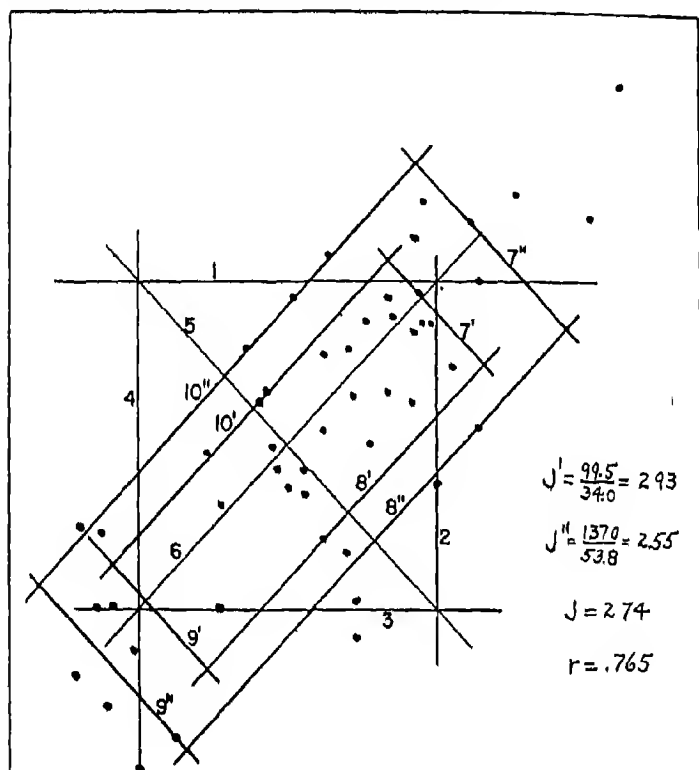


FIG. 1.

4. Move a straightedge from the left side of the scattergraph, keeping it parallel to the y -axis, until 16% of the plotted points show to the left of the straightedge. Through the latest point to appear draw a line parallel to the y -axis.

5, 6. Draw the diagonals of the rectangle formed by lines 1, 2, 3, 4.

7. Move a straightedge from the upper right corner of the scattergraph, keeping it parallel to diagonal 5, until 8% of the

plotted points show beyond the straightedge. Through the latest point to appear draw a line parallel to diagonal 5. Move the straightedge in until 16% of the plotted points show beyond the straightedge. Through the latest point to appear draw a line parallel to diagonal 5.

8. Move a straightedge from the lower right corner of the scattergraph, keeping it parallel to diagonal 6, until 8% of the plotted points show beyond the straightedge. Through the latest point to appear draw a line parallel to diagonal 6. Move the straightedge in until 16% of the plotted points show beyond the straightedge. Through the latest point to appear draw a line parallel to diagonal 6.

9. Move a straightedge from the lower left corner of the scattergraph, keeping it parallel to diagonal 5, until 8% of the plotted points show beyond the straightedge. Through the latest point to appear draw a line parallel to diagonal 5. Move the straightedge in until 16% of the plotted points show beyond the straightedge. Through the latest point to appear draw a line parallel to diagonal 5.

10. Move a straightedge from the upper left corner of the scattergraph, keeping it parallel to diagonal 6, until 8% of the plotted points show beyond the straightedge. Through the latest point to appear draw a line parallel to diagonal 6. Move the straightedge in until 16% of the plotted points show

TABLE 1
Conversion of J to r

<i>J</i>	<i>r</i>	<i>z*</i>	<i>J</i>	<i>r</i>	<i>z*</i>	<i>J</i>	<i>r</i>	<i>z*</i>	<i>J</i>	<i>r</i>	<i>z*</i>
1.0	.000	000	2.3	.682	833	3.6	.857	1.281	4.9	.920	1.589
1.1	.095	095	2.4	.704	876	3.7	.864	1.308	5.0	.923	1.609
1.2	.180	.182	2.5	.725	.916	3.8	.870	1.335	5.1	.926	1.629
1.3	.256	.262	2.6	.743	.956	3.9	.877	1.361	5.2	.929	1.649
1.4	.324	.337	2.7	.759	.993	4.0	.883	1.386	5.3	.931	1.668
1.5	.385	.406	2.8	.774	1.030	4.1	.888	1.411	5.4	.934	1.686
1.6	.438	.470	2.9	.788	1.065	4.2	.893	1.435	5.5	.936	1.705
1.7	.486	.531	3.0	.800	1.099	4.3	.898	1.459	5.6	.938	1.723
1.8	.529	.588	3.1	.811	1.131	4.4	.902	1.482	5.7	.940	1.741
1.9	.566	.642	3.2	.822	1.163	4.5	.906	1.504	5.8	.942	1.758
2.0	.600	.693	3.3	.832	1.194	4.6	.910	1.526	5.9	.944	1.775
2.1	.630	.742	3.4	.841	1.224	4.7	.913	1.548	6.0	.946	1.792
2.2	.658	.789	3.5	.849	1.253	4.8	.917	1.569			

* See R. A. Fisher's "Statistical Methods for Research Workers," pp. 202-215

$$\sigma_d^2 = \frac{1}{N} \sum d_i^2 = \frac{1}{N} \sum \left(y - \frac{\sigma_y}{\sigma_x} x \right)^2 \cos^2 \alpha$$

$$= (1-r) \frac{2 \sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2}$$

$$J'^2 = \frac{\sigma_x^2}{\sigma_d^2} = \frac{1+r}{1-r} \quad J' = \sqrt{\frac{1+r}{1-r}} \quad r = \frac{J'^2 - 1}{J'^2 + 1}$$

$$(R. A. Fisher's) Z = \frac{1}{2} \log_e \frac{1+r}{1-r} = \log_e J'$$

(The use of J'' (8%) has been added to the original method because the mean of J' and J'' empirically gives better results than J' alone)

Empirical Tests

A symmetrical correlation array of 1000 pairs was developed from two normal distributions. For this, the computed and graphic coefficients were virtually identical.

Computed r .764

Graphic r .763

The pairs were thoroughly shuffled and dealt out into packs of 50 each. Each pack of 50 pairs was plotted on a separate scattergraph and the coefficient of correlation determined

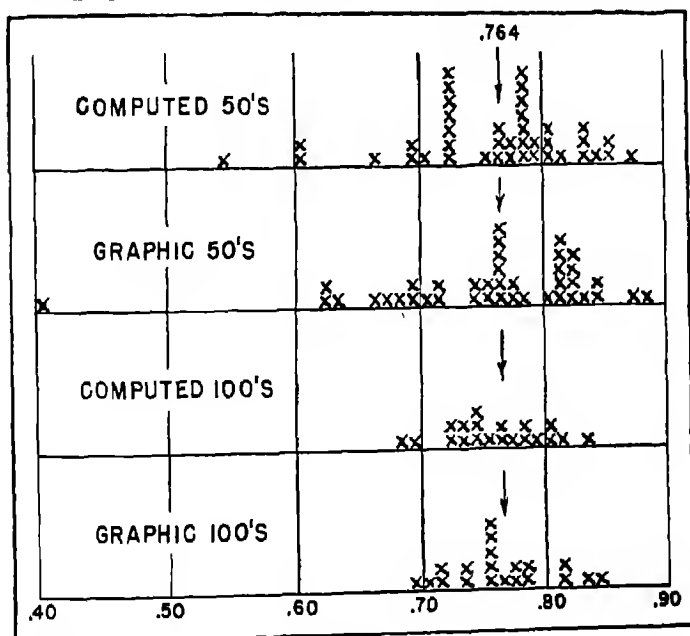


FIG. 3.

graphically and by computation. The pairs were again thoroughly shuffled and dealt out into packs of 50 each, and the graphic and computed correlations determined for each set. The same process was repeated to obtain 20 sets of 100 each.

Figure 3 shows the frequency distributions of computed and graphic coefficients for the forty sets of 50 pairs and the twenty sets of 100 pairs. Except for one low erratic value the graphic and the computed distributions are comparable. Table 2 shows

TABLE 2
Standard Deviations with One Erratic Value Omitted

	50 pairs	100 pairs
Theoretical059	.042
Computed061	.040
Graphic064	.040

the standard deviations measured from the 'r' of the whole population with one low erratic value omitted. The standard deviations for graphic and computed values are virtually identical.

In the original use of the graphic method only the lines formed by counting in 16% toward the diagonals were used. In these tests 8% and 24% were also tried. Table 3 shows com-

TABLE 3
Comparative Standard Deviations

	50 pairs	100 pairs
Mean 8% and 16%064	.040
16% alone079	.050
Mean 8%, 16%, and 24%067	..

parative standard deviations for 16% alone, for the mean of 8% and 16% ratios, and for the mean of 8%, 16%, and 24% ratios. The described method (mean of 8% and 16% ratios) is clearly better than 16% alone and is as good as the mean of 8%, 16%, and 24%, which requires much additional work.

The standard deviation of the differences between graphic and computed values is .037 for the 50-pair coefficients and also for the 100-pair coefficients. This is less than the standard

deviations of the coefficients themselves, showing that graphic and computed values were very much alike in the arrays used in these tests.

There is one type of scattergraph, however, where the graphic coefficient is bound to be quite different. That is the scattergraph having a generally symmetrical array but with a few wildly deviant cases off in one corner. In such a case, the computed coefficient may be greatly disturbed by the few deviant cases. The graphic coefficient will not be affected. In this instance, the graphic coefficient probably yields a better measure of the true relationship in the whole population from which the sample was drawn.

MEASUREMENT NEWS*

Papers relating to the field of measurement constituted more than half of those presented at the meeting of the Military Division of the American Psychological Association at the University of Maryland on November 27 and 28. The program included the following papers directly concerned with the field:

"Equivalences between Army and Civilian Tests", C. P. Sparks.

"The Army General Classification Test", Staff, Personnel Research Section, Classification and Replacement Branch, AGO, read by R. H. Bittner.

"Correction for Restricted Range", E. G. Brundage.

"The Objective Measurement of Flying Skill", A. C. Tucker.

"The Selection of Marine Officer Candidate", S. B. Williams.

"Surveys of Opinions on Training", C. R. Pace and D. L. Gibson.

"Scale and Intensity Analysis for Attitude, Opinion and Achievement", Louis Guttman.

"The Form of Items and the Distributions of False Positive Scores on a Neurotic Inventory", W. A. Owens

"War Weariness and Morale in Air Groups", J. G. Darley.

"Morale Surveys in the Army", Carl Hovland.

"The Criterion in Army Personnel Research", Staff, Personnel Research Section, Classification and Replacement Branch, AGO, read by E. D. Sisson.

"The Nominating Technique", C. L. Vaughn

"The Use of Order-of-Merit Rankings as a Criterion of Shipboard Performance of Enlisted Personnel", H. P. Bechtoldt, D. B. Stuit, and J. W. Haucker.

"Criteria of Air Crew Proficiency in Operational Training", L. B. Ward.

"Selection of Army Officers", M. W. Richardson.

"The Significance of Case History Items as Detectors of Potential Naval Delinquent", H. F. Hunt and Nathan Goldman.

"Assessment of the Whole Person: Procedures used in Testing the Suitability of 5,500 OSS Recruits", H. A. Murray.

"Test Procedures for the Psychiatric Screening of Naval Personnel: Some Problems in Method", Milton Wexler.

"Development of an Interview for Selection Purposes", Staff, Personnel Research Section, Classification and Replacement Branch, AGO, read by E. A. Rundquist.

* Readers are invited to send notes for this section to the Editor, EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 917 Fifteenth Street, NW, Washington 5, D C

The papers presented at the meeting and the discussion concerning them are to be published by the University of Maryland. Copies of the proceedings may be ordered from Professor G. A. Kelley, Department of Psychology, The University of Maryland, College Park, Maryland.

The preparation of a comprehensive handbook on educational measurement has been undertaken under the sponsorship of the Committee on Measurement and Guidance of The American Council on Education. The board of editors is composed of W. W. Cook, John C. Flanagan, E. F. Lindquist, chairman, Irving Lorge, T. R. McConnell, Philip J. Rulon, Donald J. Shank, John M. Stalnaker, Ralph Tyler, Kenneth Vaughn, and Ben D. Wood, with the chairman serving as editor-in-chief. Each of the twenty chapters is to be prepared by a specialist in the particular field covered, assisted by several collaborators. A production schedule has been outlined calling for sending final copy to the printers July 1, 1947.

An Evaluation Service Center has been established at Syracuse University, with Professor Maurice E. Troyer as director. The purposes of the Center are (1) to help faculty members in their effort to improve appraisal of student progress, (2) to assist faculty members in constructing tests, (3) to make analyses of tests, (4) to keep and make available to staff members an up-to-date file of sample published and unpublished tests in the various subject areas, (5) to keep up-to-date and make available to staff members a library of references on problems and procedures of measurement and evaluation in higher education, (6) to conduct seminars in problems of evaluation for staff members, departmental assistants and scholars interested in systematic and comprehensive study of test construction and interpretation, (7) to assist with research, (8) to encourage study and publication, by faculty members, of new and better methods of appraisal and instruction, and (9) to serve, through the staff of the Center and other members of the University Faculty on a fee basis, as consultant to other colleges and universities in the area on problems of appraisal and instruction.

The firm of Richardson, Bellows, Henry and Co., Inc., has been established at 56 Beacon Street, New York, to conduct surveys and research on problems of selection, placement, training, and employee morale for business and industrial concerns. It will do job analyses from the standpoint of qualification requirements and training needs; design application blanks, recommendation blanks, controlled interviewing procedures, aptitude tests, information tests, interest and personality tests, and training outlines and manuals; develop merit rating systems and systems for combining scores (by means of the usual multiple-correlation validation studies); make clinical appraisals of executive personnel; conduct attitude surveys; design personnel record systems and personnel statistical reporting systems;

and make over-all surveys of personnel programs. The membership consists of Roger M. Bellows, Francis F. Bradshaw (President), Edward E. Cureton (Secretary-Treasurer), Douglas H. Fryer, Edwin R. Henry, Hermann H. Remmers, Marion W. Richardson (Chairman of the Board of Directors), Carroll L. Shartle, and Robert J. Wherry (Vice President).

A counseling center has been established at the University of Chicago under the jurisdiction of the Dean of Students, Carl R. Rogers, to provide the services enumerated below. The volume of these services will be governed in accordance with the best interests of a sound program of professional education and research, carried out in cooperation with various interested departments and schools.

Services

1. To provide adjustment counseling to students, veterans, industrial workers, and other individuals and groups.
 2. To provide a diagnostic service, using tests, interviews, and other techniques.
 3. To refer individuals to appropriate University services and agencies.
 4. To assist in the coordination of specialized counseling functions on the campus.
 5. To promote the development of in-service training programs with those groups interested in improving their counseling skills.
-

William W. Blaesser, Treasurer of the American College Personnel Association and formerly of the University of Wisconsin, is now Assistant Dean of Students and Director of the Counseling Center at the University of Chicago.

Lieutenant Colonel J. P. Guilford has returned to the position of Professor of Psychology at the University of Southern California after almost four years with the Army Air Forces. In his last assignment he was Chief of the Department of Records and Analysis of the AAF School of Aviation Medicine, Randolph Field. The Department of Records and Analysis fell heir to the accumulated answer sheets, card records and data from all the examining and research units in aircrew classification. Besides turning out a number of final reports the organization completed the writing of a 29-chapter volume on *Printed Aircrew Classification Tests*, one of 15 volumes which are scheduled to be written to report the results of the AAF Psychological Research Program.

Dr. Harold C. Taylor has been appointed director of the W. E. Upjohn Institute for Community Research of Kalamazoo, Michigan. One of the major objectives of the Institute is to investigate the "suitability of opportunity for gainful employment: its relationship

to the aptitudes, skills and interests of people; and its relationship to the satisfactions, monetary and otherwise, which people desire to obtain from their jobs." Dr. Charles C. Gibbons is leaving his position as Director of Personnel Research of the Owens-Illinois Glass Company to become Assistant Director of the Institute.

Professor John M. Stalnaker has left his position with the College Entrance Examination Board and Princeton University to become Dean of Students and Professor of Psychology at Stanford University. Mr. Henry Chauncey has been appointed Associate Secretary and Professor Harold O. Gulliksen has been appointed Research Secretary of the College Entrance Examination Board.

Lieutenant Commander D. D. Feder has returned to his civilian position as Executive Officer and Supervisor of the Illinois State Civil Service Commission. His last billet in the Navy was that of Officer-in-Charge, Radio Material Unit, Test and Research Division, Bureau of Personnel.

Lieutenant Colonel Paul Horst has left the Army to return to the Procter and Gamble Company, where he is now Director of Personnel Research.

THE CONTRIBUTORS

Henry S. Dyer—Ed.D., Harvard University, 1941. Test Technician, National Clerical Ability Tests, 1938-39. Research Associate, the Graduate Record Examination Project, Carnegie Foundation for the Advancement of Teaching, 1939-41. Assistant Professor of Psychology, Allegheny College, 1941-42. Assistant Dean of Harvard College, 1942-45. Assistant to the Dean of the Faculty of Arts and Sciences, Director of the Office of Tests, Harvard University, 1945-. Member, Psychometric Society. Associate Member, American Association of University Professors.

Constance Lovell—Ph D., University of Southern California, 1942. Assistant, Psychological Laboratory, University of Southern California, 1937-42. Instructor in Psychology, University of Southern California, 1942-. Author of articles on measurement and crime. Associate Member, American Psychological Association.

Marjorie Fiske—M.A., Columbia University, 1938. Research and Administrative Assistant, Princeton Office of Radio Research, 1937-38. Research Associate, Market Analysts, Inc., 1938-39. Assistant Director, Field Service Division, National Federation of Business and Professional Women, 1939-41. Research Director, National Federation of Business and Professional Women, 1941-43. Senior Associate, Bureau of Applied Social Research, Columbia University, 1943-. Author of articles on research techniques. Associate Member, American Psychological Association.

Paul F. Lazarsfeld—Ph D., University of Vienna, 1925. Director, Division of Applied Psychology, University of Vienna, 1927-33. Rockefeller Fellow, 1933-35. Director, Office of Radio Research, 1935-. Director, Bureau of Applied Social Research, Columbia University, 1944-. Associate Professor of Sociology, Columbia University, 1943-. Author of *Radio and the Printed Page*, *Radio Research*, 1941, *Radio Research*, 1942-43, *The People's Choice*.¹ Author of numerous technical articles. Member, American Marketing Association, American Psychological Association, American Sociological Association, American Statistical Association. Fellow, American Association for Applied Psychology. Consultant to the Office of War Information, the War Department and the War Production Board.

¹ All of the works listed were published by Duell, Sloan and Pearce, New York, the dates of publication being respectively 1940, 1941, 1944, 1945.

Charles I. Mosier—Ph.D., University of Chicago, 1937. Instructor in Psychology and Vocational Guidance Counselor, University of Florida, 1933–36. Assistant Professor of Psychology, University of Florida, 1937–39. Acting University Examiner, University of Florida, 1938. Assistant Examiner, Sloan Research Project, 1940–41. Personnel Research Technician, State Technical Advisory Service, Social Security Board, 1941; Chief of Position Classification, 1942; Chief of Personnel Methods and Standards, 1943–44; Chief of Research and Test Construction, 1945–. Author of numerous articles in *Psychometrika*, *Psychological Review*, *Journal of Educational Psychology*, and others. Associate Member, American Psychological Association. Member, Psychometric Society, Southern Regional Committee of the Social Science Research Council. Member of the editorial boards of *Psychometrika* and EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT.

Helen G. Price—M.A., Stanford University, 1931. Special Research Assistant, Stanford University, 1930–32. Research Assistant, Employment Stabilization Research Institute, 1932–33. Training Supervisor, Ohio State Employment Service and Ohio Bureau of Unemployment Compensation, 1936–41. Personnel Research Technician, Social Security Board, 1941–. Author (with others) of *A Manual of Selected Occupational Tests for Use in Public Employment Offices*, Bulletin of Employment Stabilization Research Institute, Vol. II, No. 3, University of Minnesota, 1933. Associate Member, American Psychological Association.

Doncaster G. Humm—Ph.D., University of Southern California, 1932; Sc.D., Bucknell University, 1945. Counselor, Sentous Junior High School, Los Angeles, 1921–32. Psychologist, Los Angeles Diagnostic Clinic for Neuro-Psychiatry and Psychology, 1931–34. Lecturer, College of Medical Evangelists, 1929–31. Owner, Doncaster G. Humm Personnel Service. Author of articles on personality, temperament and clinical method. Co-author (with Guy W. Wadsworth, Jr.) of the *Humm-Wadsworth Temperament Scale* Specialist in the application of clinical psychology to personnel work. Member, National Institute of Industrial Psychology (Great Britain), American Mathematical Society, American Statistical Association, Society for the Advancement of Management.

J. R. Wittenborn—Ph.D., University of Illinois, 1942. Assistant Instructor in Psychology, University of Illinois, 1939–42. Research and Psychometric Assistant, Personnel Bureau, University of Illinois, 1939–42. Instructor in Psychology and Assistant Clinical Psychologist in the Department of University Health, Yale University, 1942–44. Assistant Professor of Psychology and Clinical Psychologist in the Department of University Health, Yale University, 1944–. Author of articles on mental measurements and student efficiency. Associate Member, American Psychological Association. Member, American Statistical Association. Psychological and Statistical Consultant

for a medical research project connected with the Office of Scientific Research and Development.

William M. Davidson—M.A., Indiana University, 1943. Classification Officer, Adjutant General Department, United States Army, 1943— . Associate Member, American Psychological Association.

John B. Carroll—Ph.D., University of Minnesota, 1941. Instructor in Psychology and Education, Mount Holyoke, 1940-42. Instructor in Psychology, Indiana University, 1942-43. Lecturer in Psychology, University of Chicago, 1943-44. Naval Reserve Officer attached to the Aviation Psychology Branch, Division of Aviation Medicine, Bureau of Medicine and Surgery, Navy Department, 1944— . Author of articles on factor analysis, test theory, and the psychology of language. Associate Member, American Psychological Association.

Irwin August Berg—Ph.D., University of Michigan, 1942. Personnel Counselor, Western Electric Company, 1936-39. Clinical Assistant and Teaching Fellow, University of Michigan, 1939-42. Psychologist, State Prison of Southern Michigan, summer 1942. Assistant Professor of Psychology and Clinical Counselor, University of Illinois, 1942— . Author of technical articles on criminology, personnel, and tests. Member, American Psychological Association, American Association for the Advancement of Science, American College Personnel Association, Association of Midwestern College Psychiatrists and Clinical Psychologists.

(Mrs.) Graham Johnson—M.A., Syracuse University, 1941. Assistant Psychometrist, University of Illinois, 1943-45.

Robert Peter Larsen—Ph.D., University of Iowa, 1938. Director, Iowa Reading Clinic, 1936-38. Assistant Professor of Psychology and Clinical Counselor, University of Illinois, 1938— . Author of technical articles on reading and study habits. Member, American Psychological Association, American Association for Applied Psychology, American College Personnel Association, Midwestern Association of College Psychiatrists and Clinical Psychologists.

William Leroy Jenkins—Ph.D., University of Michigan, 1936. Instructor, Assistant Professor, Lehigh University, 1935-43. Research Associate, University of California Division of War Research, 1943-44. Supervisor, Training Aids, Columbia University Division of War Research, Submarine Training Section, 1944-45. Associate Professor of Psychology, Lehigh University, 1946— . Author of articles on cutaneous sensitivity. Member, American Psychological Association.

MEASUREMENT ABSTRACTS*

Berdie, Lt. R. F. "Range of Interests." *Journal of Applied Psychology*, XXIX (1945), 268-281.

An interest scale based on 22 items was found to differentiate clearly between recruits who could be expected to adjust to military training and those who could not. As against the orally presented list, the printed list proved of greater convenience and objectivity. Age and educational factors must be taken into account in any analysis of the results. Supplementing the psychiatric screening technique and the interview, this range of interests test offers a satisfactory method of predicting adjustment. *Vernon S. Tracht.*

Challman, Robert C. "The Validity of the Harrower-Erickson Multiple Choice Test as a Screening Device" *Journal of Psychology*, XX (1945), 41-48

The *Harrower-Erickson Multiple Choice Test* was designed as a selective device for military use. The procedure consists of offering subjects 10 choices for each of the 10 Rorschach cards. Half of the choices are considered representative of individuals suffering from mental abnormalities. If the subject considers none of the choices applicable, he is advised to submit an alternate. The critical score is based on 4, 5, or 6 abnormal responses, depending upon the degree of selectivity desired. Three methods of scoring were suggested. In Method I all alternates are classed as abnormal; in Method II alternates are scored abnormal only when characterized by poor form or when bizarre in content; in Method III abnormal and alternate responses are weighted. Harrower-Erickson found the procedure valuable as a screening device. However, later studies, including the one described in this article, do not indicate a sufficiently sharp distinction between the responses of the normal and the abnormal to warrant the acceptance of the method as more than an auxiliary to be used with a personality inventory. *Helen Heath.*

Forlano, G. and Kirkpatrick, F. H. "Intelligence and Adjustment Measurements in the Selection of Radio Tube Mounters" *Journal of Applied Psychology*, XXIX (1945), 257-261

This study was concerned with the problem of effectiveness of intelligence and adjustment tests in bringing about increased worker efficiency in radio tube mounting jobs. Subjects in the experiment were 20 female tube mounters. Tests used were (1) the *Otis Self-Administering Test of Mental Ability, Form B*, (2) the social scale of the *Bell Adjustment Inventory*; (3) the alienation scale of the *Washburne Social Adjustment Inventory*. The criterion used for the experiment was ratings by the supervisor in charge of the group. It is concluded that low intelligence scores tend to indicate poorer workers but average scores or above do not discriminate between "good" and "fair" workers, while scores in social adjustment do differentiate "good" and "fair" workers. A composite of intelligence and personality scores is therefore effective in predicting success of new tube mounters. *Frances Smith.*

Geil, George A. "A Clinically Useful Abbreviated Wechsler-Bellevue Scale." *Journal of Psychology*, XX (1945), 101-108.

Selection of a shortened form of the Bellevue full scale, which would meet requirements of the clinical situation for time economy, accuracy of intelligence

* Edited by Forrest A. Kingsbury.

rating, and diagnostic screening capacity, was made on the basis of analysis of test records of a group of 250 unselected cases examined by the Wechsler-Bellevue full scale at the Medical Center for Federal Prisoners, Springfield, Mo. Mean weighted scores for each of the 10 subtests of the scale, and mean total full scale weighted subtest score were determined, and mean total weighted subtest scores for trial combinations of selected subtests were then computed. Trial combinations were retained which showed close alignment of mean total weighted subtest scores with that indicated for the full scale, and were tested for accuracy by comparison of calculated IQ scores on the combinations for each of the 250 cases with the actual full scale IQ scores. An abbreviated scale composed of the tests of Comprehension, Similarities, Digits, and Block Design, shows a correlation of $.966 \pm .003$ with the IQ's of the full scale, and is found to be reliable as a screening instrument. *Frances Smith*

Gulliksen, Harold. "The Relation of Item Difficulty and Inter-Item Correlation to Test Variance and Reliability." *Psychometrika*, X (1945), 79-91

Under assumptions that will hold for the usual test situation, it is proved that test reliability and variance increase (a) as the average inter-item correlation increases, and (b) as the variance of the item difficulty distribution decreases. As the average item variance increases, the test variance will increase, but the test reliability will not be affected. (It is noted that as the average item variance increases, the average item difficulty approaches .50). In this development, no account is taken of the effect of chance success, or the possible effect on student attitude of different item difficulty distributions. In order to maximize the reliability and variance of a test, the items should have high intercorrelations, all items should be of the same difficulty level, and the level should be as near to 50% as possible. (Courtesy *Psychometrika*.)

Hildreth, Lt. H. M. "Single-Item Tests for Psychometric Screening" *Journal of Applied Psychology*, XXIX (1945), 262-267

This describes how a series of 10 Single-Item Tests, selected from 30 experimental items from well-known mental tests by a special scoring technique, were devised for use in the Navy's psychometric screening program. A successful response to any one of these tests indicated mental ability above the minimum—M.A. of 11 years—arbitrarily selected by naval officials. Taking from 1 to 20 minutes per man to administer, and not requiring optimal testing conditions, they were standardized on 1500 cases; and, in a 3 months' trial period, exhibited an error in the selection of recruits of only 1/10 of 1 per cent. *Vernon S. Tracht*.

Hult, Esther "Study of Achievement in Educational Psychology." *The Journal of Experimental Education*, XIII (1945), 174-190

The purpose of this study was to find the relationship between practice teaching success and measures of ability considered prerequisite for this success as determined by various tests. The criteria were (1) practice teaching marks, and (2) ratings by supervisors. No significant relationships were found between practice teaching marks, general knowledge and mental ability. The multiple correlation between the several factors and the criterion was not high enough for individual prediction but there was a significant relationship between success in teaching and the grade point average. Shortly before the end of the semester the practice teachers rated their theory course and teacher, and practice course and teacher. Of the two courses, they tended to rate their practice course higher, and of the two teachers, their practice teacher lower. *Howard M. Schuman*.

Johnson, Palmer O. and Tsao, Fei. "Factorial Design and Covariance in the Study of Individual Educational Development." *Psychometrika*, X (1945), 133-162

This is the report of the application of the principles of factorial design to an investigation of individual educational development. The specific type of factorial design formulated was a $2 \times 3 \times 3 \times 3$ arrangement, that is, the effect of sex, grade location, scholastic standing, and individual order, singly and in all possible combinations.

nations was studied in relation to educational development as measured by the *Iowa Tests of Educational Development*. An application of the covariance method was introduced which resulted in increased precision of this type of experimental design by significantly reducing experimental error. The two concomitant measures used to increase the sensitiveness of the experiment were initial status of individual development and mental age. Without these statistical controls all main effects and two first-order interactions would have been accepted as significant. With their use only sex (doubtful), scholastic standing, and individual order demonstrated significant effects. The chief beauty of the analysis of variance and covariance as an integral part of a self-contained experiment is demonstrated in the complete single analysis of the data. The statistical utilization of the experimental results has also been developed for purposes of estimation and prediction. The mathematical statistician is being continuously required to develop and analyze experimental designs of increasing complexity since the introduction of the analysis of variance and covariance. The mathematical formulation and solution of the problem of this investigation is carried out. The methods illustrated and explained in this study, and modifications and extensions of them are capable of very wide application. The general principles can be used to various degrees and in a number of ways. (Courtesy *Psychometrika*.)

Kaitz, Hyman B. "A Note on Reliability." *Psychometrika*, X (1945), 127-131

A formula for internal consistency reliability is developed within the framework of the analysis of variance. The test items are assumed to be homogeneous, but may have any weights. Data needed for computation are the student test scores and the total number of items answered so as to have the same weight. It is shown that this formula reduces to the Kuder-Richardson for item weights of one and zero. Some empirical validation is offered. (Courtesy *Psychometrika*.)

Martin, Howard G. "The Construction of the Guilford-Martin Inventory of Factors G-A-M-I-N." *Journal of Applied Psychology*, XXIX (1945), 298-300

Factor analyses have isolated five new temperament traits, (G) general pressure for overt activity, (A) social ascendancy, (M) masculinity of attitudes and interests, (I) self-confidence, (N) lack of nervous tension. Over 300 items, answered "Yes," "No," or "?" were administered to 250 men and 250 women, all college students between the ages of 19 and 30. Items, shown by factor and item analyses to have heavy loadings in a trait, were used on the preliminary scoring key. Four hundred tests taken by 200 men and 200 women were scored with this preliminary key, and the highest 100 and lowest 100 cases were used as criterion groups for factors G, A, I, and N. Factor M was based on scores of the highest 100 males and the lowest 100 females. Split-half reliability on the five traits ranged from .85 to .91. *Howard M. Schuman.*

Newman, Joseph. "The Prediction of Shopwork Performance in an Adult Rehabilitation Program. The Kent-Shakow Industrial Formboard Series." *Psychological Record*, V (1945), 343-352.

An investigation of the value of the K-S Formboard Series for predicting performance in shopwork was conducted by means of a study extending over two years, with results based on data obtained from 111 male patients in a New York sanatorium who took part in a rehabilitation program. Subjects were given the K-S Formboard before assignment to the wood-working shop; shopwork progress was determined by means of rating scales and subjects were also ranked in ascending order according to total time score on the K-S Formboard. Formboard results were studied to determine their relationship to shopwork ratings. It is concluded that the K-S Formboard is of value for predicting performance in shopwork in an adult re-education program. A differentiating score for the Formboard is a total time of 25 minutes or less. A correlation coefficient (tetrachoric) of .76 was found between shop ratings in accuracy, speed, and constructive thinking, and total time scores. *Frances Smith.*

Peel, E. A. "On Identifying Aesthetic Types" *British Journal of Psychology*, XXXV (1945), 61-69

This paper outlines a method for estimating aesthetic preference with reference to artistic quality rather than to temperamental traits involved. Items were arranged by expert judges according to aesthetic criteria and the subjects' orders of aesthetic preference were compared with these criteria by means of correlation. The correlations were then analyzed into factors characterizing the group of subjects and the criteria. Three different matrices of correlation coefficients were obtained, correlations between orders of "liking" for the subjects, correlations between the orders of "liking" and the criteria, and correlations between the criteria. *Frances Smith*

Rabin, Albert I. "The Use of the Wechsler-Bellevue Scales with Normal and Abnormal Persons" *Psychological Bulletin*, XLII (1945), 410-422

Findings of various investigators employing the *Wechsler-Bellevue Scale* are coordinated and summarized and suggestions are offered for future treatment of the data rapidly being accumulated. Correlations between this scale and other tests of intelligence and achievement are reported, and the diagnostic effectiveness of the scale is substantiated. The scale is demonstrated to be effective in group pattern-analysis because of the functional unity of its subtests; and the possibility of achieving a diagnostic tool in individual cases through more effective control of major factors involved is discussed. Retest data with clinical material and miscellaneous studies are reported and the need for long-range retest studies and further investigation by the method of factor analysis in different clinical groups and at different age levels is emphasized. *Frances Smith*

Sarason, E. and Sarason, S. "A Problem in Diagnosing Feeble-Mindedness" *Journal of Abnormal and Social Psychology*, XL (1945), 323-329.

Criteria are formulated by means of which a clinical psychological report may be judged in diagnosis of mental deficiency. These criteria are (1) inclusion in the psychological examination of several measures of intelligence of the individual type of test, (2) use of projective techniques to clarify the relation between intelligence and personality, (3) internal analysis of each test, (4) interpretation of test functioning as part of a continuous behavioral sequence; (5) integration of information obtained from the various tests. Need for care in acceptance of numerical scores in doubtful and near-borderline cases is illustrated by presentation of the complete psychological report of a particular case. *Frances Smith*.

Shakow, D., Rodnick, E. H., and Lebeaux, T. "A Psychological Study of a Schizophrenic: Exemplification of a Method" *Journal of Abnormal and Social Psychology*, XL (1945), 154-174.

A collection of 8 psychological devices: (1) *Stanford-Binet* or *Wechsler-Bellevue Intelligence Scale*, (2) *Rorschach Test*, (3) *Association Test*, (4) *Pinboard Aspiration Test*, (5) *Thematic Apperception Test*, (6) *Targetball-Thematic Test*, (7) *Pursuitmeter-Stress Test*; and (8) *Picture-Frustration Test* were employed as one aspect of a comprehensive study of neuropsychiatric patients who had been in service in the armed forces. The specific aim of the psychological analysis was to construct individual profiles and also to differentiate patient groups from each other and from normal groups. One case was reviewed in detail. *Helen Heath*

Thurstone, L. L. "A Multiple Group Method of Factoring the Correlation Matrix" *Psychometrika*, X (1945), 73-78.

There are a number of methods of factoring the correlation matrix which require the calculation of a table of residual correlations after each factor has been extracted. This is perhaps the most laborious part of factoring. The method to be described here avoids the computation of residuals after each factor has been computed. Since the method turns on the selection of a set of constellations or clusters of test vectors, it will be called a *multiple-group method of factoring*. The method can be used for extracting one factor at a time if that is desired but it will be considered here for the

more interesting case in which a number of constellations are selected from the correlation matrix at the start. The result of this method of factoring is a factor matrix F which satisfies the fundamental relation $FF' = R$ (Courtesy *Psychometrika*)

Walker, K. F., Staines, R. G., and Kenna, J. C. "The Influence of Scoring Methods upon Score in Motor Perseveration Tests." *British Journal of Psychology, General Section*, XXXV (1945), 51-60.

Spearman based his theory of mental inertia on the results of motor perseveration tests. Certain features in the construction and scoring of these tests invalidate these results. The two types of these tests are (1) creative effort, and (2) alternation. The five methods of scoring these tests are: (a) $X - Y$, (b) X/Y , (c) $X - X/X$, (d) $X + Y/2XY$, and (e) E/A where X and Y are two interfering tasks, E is the expected score, and A the actual score. Methods (a), (b), and (c) correlate highly with each other but low or negatively with methods (d) and (e). Spearman's general interference factor, found when method (d) is used, disappears when method (e) is used. When initial differences in ease of performing the two activities are great, the creative effort test using method (d) does not measure difficulty in alternation but only the difference in ease of performance. The initial difference in ease of performance is not related to ease of alternation of the two activities. *Howard M. Schuman*.

Werner, Heinz (with the collaboration of Doris Carrison). "Perceptual Behavior of Brain-Injured, Mentally Defective Children: An Experimental Study by Means of the Rorschach Technique." *Genetic Psychology Monographs*, XXXI (1945), 53-110.

Experimental analysis of perceptual and conceptual behavior of brain-injured and non-brain-injured subnormal children of comparable mental ages was conducted by means of the Rorschach technique. Significant differences in response were found, behavior of brain-injured children being characterized by disintegrative tendencies, forced responsiveness to sensory stimulation, lack of affective motor-control, lack of associational control, meticulousity and perseverations. Interpretation of these responses is made in the light of previous studies including experiments with similarly formed groups of children, work on the Rorschach test with brain-injured adults, and general studies of responses to the Rorschach test. Characteristic clusters of behavior traits of brain-injured children are deduced from this analysis. *Frances Smith*.
